



## **VMStools: Open-source software for the processing, analysis and visualization of fisheries logbook and VMS data**

**Hintzen, Niels T.; Bastardie, Francois; Beare, Doug; Piet, Gerjan J.; Ulrich, Clara; Deporte, Nicolas; Egekvist, Josefine; Degel, Henrik**

*Published in:*  
Fisheries Research

*Link to article, DOI:*  
[10.1016/j.fishres.2011.11.007](https://doi.org/10.1016/j.fishres.2011.11.007)

*Publication date:*  
2012

[Link back to DTU Orbit](#)

*Citation (APA):*  
Hintzen, N. T., Bastardie, F., Beare, D., Piet, G. J., Ulrich, C., Deporte, N., Egekvist, J., & Degel, H. (2012). VMStools: Open-source software for the processing, analysis and visualization of fisheries logbook and VMS data. *Fisheries Research*, 115-116, 31-43. <https://doi.org/10.1016/j.fishres.2011.11.007>

---

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## **VMStools: Open-source software for the processing, analysis and visualisation of fisheries logbook and VMS data**

Niels T. Hintzen<sup>1\*</sup>, Francois Bastardie<sup>2</sup>, Doug Beare<sup>1</sup>, Gerjan Piet<sup>1</sup>, Clara Ulrich<sup>2</sup>, Nicolas Deporte<sup>3</sup>, Josefine Egekvist<sup>2</sup>, Henrik Degel<sup>2</sup>

<sup>1</sup>IMARES, part of Wageningen UR, Institute for Marine Resources and Ecosystem Studies, PO Box 68, 1970 AB IJmuiden, The Netherlands

<sup>2</sup>DTU-Aqua, Technical University of Denmark, National Institute of Aquatic Resources, Charlottenlund Castle, DK-2920 Charlottenlund, Denmark

<sup>3</sup>IFREMER, French Research Institute for Exploration of the Sea, Brest, STH, BP 70, 29280 France

### **ABSTRACT**

**Vmstools** is a package of open-source software, build using the freeware environment R, specifically developed for the processing, analysis and visualisation of landings (logbooks) and vessel location data (VMS) from commercial fisheries. Analyses start with standardized data formats for logbook (EFLALO) and VMS (TACSAT), enabling users to conduct a variety of analyses using generic algorithms. Embedded functionality handles erroneous data point detection and removal, métier identification through the use of clustering techniques, linking logbook and VMS data together in order to distinguish fishing from other activities, provide high-resolution maps of both fishing effort and - landings, interpolate vessel tracks, calculate indicators of fishing impact as listed under the Data Collection Framework at different spatio- temporal scales. Finally data can be transformed into other existing formats, for example to populate regional databases like FishFrame. This paper describes workflow examples of these features while online material allows a head start to perform these analyses. This software incorporates state of-the art VMS and logbook analysing methods standardizing the process towards obtaining pan-European, or even worldwide indicators of fishing distribution and impact as required for spatial planning.

**Keywords:** Area based management, fishing impact, indicators, marine spatial planning, métier analyses

\*Corresponding author: tel: +31 317 489070; e-mail: [Niels.Hintzen@wur.nl](mailto:Niels.Hintzen@wur.nl)

**Article first published online: January 2013**

Please note that this is an author-produced PostPrint of the final peer-review corrected article accepted for publication. The definitive publisher-authenticated version can be accessed here:

<http://dx.doi.org/10.1016/j.fishres.2011.11.007> © 2013 Elsevier

# **vmstools: open-source software for the processing, analysis and visualization of fisheries logbook and VMS data**

Niels T. Hintzen<sup>1\*</sup>, Francois Bastardie<sup>2</sup>, Doug Beare<sup>1</sup>, Gerjan Piet<sup>1</sup>, Clara Ulrich<sup>2</sup>, Nicolas Deporte<sup>3</sup>, Josefine Egekvist<sup>2</sup>, Henrik Degel<sup>2</sup>

<sup>1</sup>IMARES, part of Wageningen UR, Institute for Marine Resources and Ecosystem Studies, PO Box 68, 1970 AB IJmuiden, The Netherlands

<sup>2</sup> DTU-Aqua, Technical University of Denmark, National Institute of Aquatic Resources, Charlottenlund Castle, DK-2920 Charlottenlund, Denmark

<sup>3</sup> IFREMER, French Research Institute for Exploration of the Sea, Brest, STH, BP 70, 29280 Plouzané, France

\*Corresponding author: tel: +31 317 489070; fax: +31 317 487326; e-mail: Niels.Hintzen@wur.nl

## **Abstract**

**Vmstools** is a package of open-source software, build using the freeware environment R, specifically developed for the processing, analysis and visualisation of landings (logbooks) and vessel location data (VMS) from commercial fisheries. Analyses start with standardised data formats for logbook (EFLALO) and VMS (TACSAT), enabling users to conduct a variety of analyses using generic algorithms. Embedded functionality handles erroneous data point detection and removal, métier identification through the use of clustering techniques, linking logbook and VMS data together in order to distinguish fishing from other activities, provide high-resolution maps of both fishing effort and - landings, interpolate vessel tracks, calculate indicators of fishing impact as listed under the Data Collection Framework at different spatio-temporal scales. Finally data can be transformed into other existing formats, for example to populate regional databases like FishFrame. This paper describes workflow examples of these features while online material allows a head start to perform these analyses. This software incorporates state-of-the art VMS and logbook analysing methods standardizing the process towards

32 obtaining pan-European, or even worldwide indicators of fishing distribution and impact  
33 as required for spatial planning.

34

## 35 **Keywords**

36 Area based management, fishing impact, indicators, marine spatial planning, métier  
37 analyses

38

## 39 **1. Introduction**

40 Growing pressures by various human activities on the marine environment and  
41 international commitments to the conservation of biodiversity or seafloor integrity (CEC,  
42 2007) have led to increased interest in marine spatial planning and in the tools required  
43 for an assessment of the impact of these pressures (Douvere and Ehler, 2009). Fishing is  
44 considered, given its widespread occurrence, to probably be the main human activity  
45 impacting the seafloor (Eastwood, 2007; Kaiser et al., 2006). Vessel Monitoring by  
46 Satellite (VMS) system data on the spatial distribution of fisheries have been collected  
47 from 2000 onwards (EC, 2002); (Piet et al., 2007), originally introduced for control  
48 purposes. However, no regional assessments on the spatial impact of international fishing  
49 activities on the seafloor have yet been conducted at appropriate scales (Piet and  
50 Quirijns, 2009). An important reason relate to concerns of confidentiality and commercial  
51 sensitivity over the use of raw VMS data, as points identify exact vessel positions.  
52 However, aggregating VMS into métiers following strict protocols should overcome this  
53 and thereby facilitating the wider exchange of data in Europe (Lee et al., 2010).  
54 However, defining métiers is in itself a difficult issue, as identifying distinct and well-  
55 defined types of fishing activities can be executed using a variety of criteria and  
56 methods, often including a part of subjectivity (ICES, 2003). And in spite of intense  
57 scientific activity in this field over the last two decades, no standardised approach for  
58 defining métiers across regions and countries has yet fully emerged (Ulrich et al., 2009).  
59 Time is pressing to deal with these technical issues, as the implementation of an  
60 ecosystem approach to fisheries management and part of the revised Common Fisheries

61 Policy, require a move towards fleet and area-based management (EC, 2008). These  
62 advances may be used to direct marine spatial planning and to reduce the pressure by  
63 human activities on the marine environment.

64

65 In European Union (EU) member states, spatial fishing information can be obtained from  
66 two main sources: the logbooks and the VMS data (EC, 2002). (i) Logbooks are the  
67 responsibility of the skipper of each vessel and have been mandatory on all commercial  
68 fishing vessels larger than 10 meters cruising in EU waters, since 1985, or when landings  
69 exceed 50kg (EC, 1993; Long and Curran, 2000). Logbook data, here referred to as the  
70 combined dataset of the fleet register, logbook data filled out by skippers and sale slips,  
71 provide information on aspects of the fishing operations (gear types used, mesh size,  
72 landings) and the physical characteristics of each vessel (vessel size, engine power). In  
73 their logbooks the fishermen must also declare the location (usually at the ICES  
74 statistical rectangle level, 1° longitude, 0.5° latitude) and the date where each landing  
75 was taken. (ii) The VMS regulations (mandatory on vessels >24m in length from 2000-  
76 2004; and >15m from 2005-2011), first introduced in January 2000, require the regular  
77 submission (via satellite) of the exact locations (longitude, latitude, speed and heading)  
78 of each vessel to a centralized database. Typically the intervals between positions or  
79 pings are one or two hours.

80

81 In the past decade, VMS analyses have mainly focused on mapping fishing effort  
82 distribution (see a review in Lee et al., 2010) and on refining the methodology for  
83 describing fishing tracks or activity (Mills et al., 2007; Hintzen et al., 2010; Vermard et  
84 al., 2010). Some recent studies have explored methods for allocating logbook catches to  
85 VMS positions (Bastardie et al., 2010b; Gerritsen and Lordan, 2011). Hence, logbook and  
86 VMS data are complementary and the coupling of logbook and VMS data has already  
87 proven powerful, also for describing the spatial distribution of marine biota habitat at a  
88 much finer spatial or temporal resolution (Bastardie et al., 2010b; Eastwood, 2007; Fock,  
89 2008; Gerritsen and Lordan, 2011; Hintzen et al., 2010; Lee et al., 2010; Mills et al.,

2007; Pedersen et al., 2009; Stelzenmuller et al., 2008; Vermard et al., 2010; Walker and Bez, 2010).

In most of the studies listed above, however, the data have been processed with *ad hoc* tools, making the analyses difficult to repeat even if the methodology is well described. The use of pre-defined (standardized) data formats in combination with standard scripts would allow various operators to perform identical analyses on similar data sources (Cagnacci and Urbano, 2008), and would therefore provide opportunities to accelerate our understanding of the marine habitat and its use (Reichman et al., 2011; Kranstauber et al. 2011). Ecopath with Ecosim (Christensen and Walters, 2004), as a standardized framework for example, has revolutionised ecosystem modelling, and its applications are found all over the world. Likewise, **FLR** (Kell et al., 2007) which is also an R 'add-on' package (R Development Core Team, 2008) has proved to be extremely useful for standardising stock assessments and management strategy evaluations (Kraak et al., 2008; Pastoors et al., 2007; Sainsbury et al., 2000). In this paper we demonstrate how the use of the **vmstools** package to jointly analyse VMS and logbook data improves our understanding of the marine habitat and its usage by fisheries.

**Vmstools** uses two standardized data formats, EFLALO (EU logbook data) and TACSAT (the VMS positions). These formats build on work done and agreements made during previous EU funded scientific projects such as TECTAC, CAFÉ and AFRAME (see Appendix A) and are well known within the International Council for the Exploration of the Sea (ICES) community. Once the data have been imported into R, a series of functions linked by scripts enable a range of tasks to be completed in a single software environment. Métiers, for example, can be identified objectively from logbook landings species compositions using multivariate and clustering techniques; fishing activity can be distinguished from other activities (i.e. vessels in harbour or steaming); logbook and VMS data can be linked (Bastardie et al., 2010b) and individual vessel tracks can be interpolated both linearly and non-linearly using Hermite spline functions (Hintzen et al., 2010). The package can furthermore be used to explore the effect of different spatial

120 (grid size) and temporal aggregations (monthly, quarterly, annually) which can be  
121 extremely important when determining fishing impact and its indicators (Piet and  
122 Quirijns, 2009), of which three of those listed under the Data Collection Framework are  
123 embedded within the package too.

124

## 125 **2. Material & Methods**

126 This software was developed during the EU funded project 'Development of tools for  
127 logbook and VMS data analysis (MARE/2008/10 lot 2)'. The open-source statistical  
128 computing environment, R, was selected because it is free, is used widely in the fisheries  
129 scientific community, and already incorporates a range of useful add-on packages  
130 capable of dealing with spatial data. A public repository has been created for hosting the  
131 development of the package, from which the latest version of ***vmstools*** can also be  
132 downloaded (<http://code.google.com/p/vmstools/>). Each program submitted to the  
133 repository must have a manual describing its use and, furthermore, must be designed to  
134 use TACSAT and EFLALO formats. (Note: from here on 'tacsat' will be used as a reference  
135 to the formatted VMS dataset while 'eflalo' is used as a reference to the formatted  
136 logbook data). Anyone with an interest in analysing such data can get involved by  
137 contacting the authors. Illustrations of the use of the tools (for example scripts, see also  
138 Appendix B) are also available on the repository. These scripts describe possible ways the  
139 R functions can be combined when analysing and coupling VMS and logbook data. The  
140 most salient points are described below (references to R functions currently incorporated  
141 within the package are given in *italics*).

142

### 143 2.1 Data

144 VMS and logbook processing using ***vmstools*** first requires eflalo and tacsat data to be  
145 created in the correct format. Each of these datasets has a pre-defined structure (see  
146 Appendix A) with certain mandatory columns. Data are imported into R from csv files  
147 (comma separated values) using *readEflalo* and *readTacsat*, which ensures that each  
148 column of data is in the correct format (internal check by *formatEflalo* and *formatTacsat*),  
149 e.g. date, character or number. To illustrate these datasets the Dutch fishing industry

150 and Ministry kindly gave the authors the permission to incorporate subsets of raw  
151 logbook and VMS data directly into the **vmstools** package. For confidentiality purposes,  
152 these data have been disguised, and noise has been added to the recorded vessel  
153 positions. These data are now thus an embedded component of the package, allowing  
154 potential users to test and demonstrate the software. These example datasets are also  
155 used to illustrate the software functioning within this paper.

156

## 157 2.2 Cleaning the data

158 In many instances, both VMS and EU logbook data contain erroneous entries. It is  
159 common to find vessel positions on land, implausibly high speeds, headings outside a  
160 compass range, and duplicate records. As advised by ICES (ICES, 2010), these errors  
161 should be removed or flagged. In addition, vessel positions lying either in harbours, or  
162 very close to harbours should be identified. The **vmstools** package distinguishes such  
163 records using standard GIS-type point-in-polygon calculations. Functions such as,  
164 *sortTacsat*, *pointOnLand* and *pointInHarbour* are examples of how this process is  
165 facilitated, supported by command line access to extensive harbour position lists and  
166 European coastline shapefiles.

167 Within this exercise, positions in the tacsat file which were outside longitudinal or  
168 latitudinal ranges (latitude > 90 or latitude < -90, longitude < -180 or longitude > 180),  
169 or had speed records of more than 20nmh<sup>-1</sup> were removed. After filtering out duplicate  
170 records, points in harbour were removed (*pointInHarbour*), as were points on land  
171 (*pointOnLand*, see figure 1, table 1). In addition to 'true' duplicates, the tacsat file was  
172 also filtered for pseudo-duplicate positions where records with intervals of less than 5  
173 minutes were removed. Spurious mesh sizes (> 150mm in the example dataset) in the  
174 eflalo dataset were also removed, while landings that were larger than approximately 30x  
175 any other landing recorded of that species were flagged. In addition, non-unique trip IDs  
176 were removed where a trip ID consisted of the combination of vessel ID, trip number,  
177 haul and ICES statistical rectangle of the landing.

178

## 179 2.3 Métier Identification



180 Fishing operations and fishing trips showing similar patterns need to be grouped by  
181 métiers according to the Data Collection Framework for the European Union (DCF, EC  
182 2008). The DCF has defined métiers according to a hierarchical structure using six nested  
183 levels: Level 1- Activity (fishing/non fishing), Level 2- Gear class (e.g. trawls, dredges),  
184 Level 3- Gear group (e.g. bottom trawls, pelagic trawls), Level 4- Gear type (e.g. Bottom  
185 otter trawl (OTB), Bottom pair trawl (PTB)), Level 5- Target assemblage based on main  
186 species type (e.g. Demersal fish, Crustaceans), Level 6- Mesh size and other selective  
187 devices. Because logbooks do not contain the information on the assemblage of targeted  
188 species as required by the definition of the métier at the level 5, a number of methods  
189 have been used in the past to identify these, typically using a variety of statistical  
190 analyses on landings profiles from logbooks data (see also the reviews by Marchal, 2008  
191 and Ulrich et al., 2009). In order to implement and compare some of the most commonly  
192 used of these methods in a generic and objective framework, a workflow has been set up  
193 to apply multivariate and clustering analyses to the logbook landings composition, in  
194 order to deduce it for each fishing operation (logbook event). The details of the  
195 methodology included in the present **vmstools** library are the subject of another paper  
196 (Deporte et al., in prep) and only the salient points need to be recapitulated here. The  
197 steps undertaken can be summarized as follows:

198 (i) First it is necessary to identify the most valuable species in the logbooks so the size of  
199 the dataset can be rendered manageable (*selectMainSpecies* and  
200 *extractTableMainSpecies*). (ii) Secondly the total inertia of the dataset must be reduced  
201 by applying the routines *getTableAfterPCA* which runs principal component analysis (PCA)  
202 followed by a selection of clustering methods: Hierarchical Agglomerative Classification  
203 (HAC, Hartigan, 1975), K-Means (Hartigan and Wong, 1979) or Clustering LARge  
204 Applications (CLARA) algorithm (Kaufman and Rousseeuw, 1990), *getMetierClusters*. The  
205 PCA aims to make the clustering process easier by reducing the amount of information  
206 comprised in the dataset to its substantial part only. Clusters group similar logbook  
207 events and are characterized by specific assemblages of different species (which are  
208 conveniently referred to as Level7, by opposition to the pre-defined Level 5 assemblages  
209 based on species type). (iii) Thirdly, a conversion of métier at this Level 7 to métier DCF

210 Level 5 categories is executed using *compareToOrdination*. This latter function also  
211 permits the comparison of these results with those obtained using alternative simple  
212 ordination methods. (iv) Finally any newly derived logbook data can be allocated into the  
213 pre-defined categories or métiers using a discriminant analysis embedded in the  
214 *predictMetier* function. To ease the whole workflow, all these sequential steps (except for  
215 the last one) have also been pooled into one single routine (*getEflaloMetierLevel7*), which  
216 reads an EFLALO dataset in and returns it together with a métier definition both at Level  
217 5 and at Level 7.

218

219 The first step of the métier identification on the example eflalo dataset consists of  
220 determining the main species, realized using the *selectMainSpecies* function.  
221 This function encompasses three methods: (i) species selection by HAC clustering, (ii)  
222 species selection by their proportion of the total catch, and (iii) species selection by their  
223 proportion of the catch of at least one logbook event. Each method gives a set of species  
224 and returns a set of main species to be included in the métier identification. After  
225 defining the main species, the original input dataset is subset to only these species  
226 (*extractTableMainSpecies*). PCA analyses are executed to reduce the dimension of the  
227 data with the function *getTableAfterPCA*. The “70% of the initial inertia” criterion was  
228 used to determine the number of axis to retain. The CLARA clustering algorithm was  
229 finally used to define the clusters.

230

## 231 2.4 Linking tacsat and eflalo data

232 By linking tacsat and eflalo data investigators can potentially explore the spatial  
233 distribution of fishing effort and landings in much greater detail than was hitherto  
234 possible (Fock, 2008; Gerritsen and Lordan, 2011; Hintzen et al., 2010; Lee et al., 2010;  
235 Piet et al., 2007). Linking tacsat and eflalo data implies that individual tacsat pings can  
236 be assigned to a particular trip as given in the eflalo dataset. This step is particularly  
237 important, as all subsequent analyses depend on the success of the linking. Linking both  
238 sources of information requires identifying common vessels identifiers (ID), and date and  
239 time limits that define the start and end of a trip or logbook event. The simplest

240 approach is to select the VMS positions that occurred between the departure and arrival  
241 dates for each trip described in the logbook data, and to assign unique trip identifiers to  
242 them. Sometimes, however, it is not possible to match tacsat records with every trip  
243 identified in the eflalo data, and in these cases the non-matching observations can be  
244 flagged with a '0' (*mergeEflalo2Tacsat*). Another, more sophisticated method available in  
245 the **vmstools** package (see Bastardie et al., 2010b), links trips by their midpoint  
246 (*mergeEflalo2Pings*).

247

## 248 2.5 Fishing activity

249 When investigating the behaviour of fishers or analysing, e.g. the impact of fishing on the  
250 seabed, it is necessary to distinguish different activities. In most instances, a distinction  
251 is made between drifting, fishing and steaming based on speed thresholds (Bastardie et  
252 al., 2010b; Fock, 2008; Rijnsdorp et al., 1998), although it has also been shown that  
253 better estimates can result if the information represented by vessel heading is utilised  
254 (Mills et al., 2007; Vermard et al., 2010). Although none of these methods will result in  
255 perfect identification of fishing behaviour, application does result in a marked  
256 improvement of our perception of spatial and temporal fishing activity and its effects on  
257 the ecosystem (Eastwood et al., 2007). Methods to identify, and therefore quantify  
258 fishing activity have been incorporated into the *segmentTacsatSpeed* (Bastardie et al.,  
259 2010b) or the *filterTacsat* functions in **vmstools**. There are many possible ways that  
260 fishing activity can be summarized. One can simply sum tacsat pings (where a ping  
261 represents the transmitted hourly or two hourly record of a vessels ID, position, speed,  
262 heading and date/time stamp), fishing time or fishing distance over any spatial  
263 compartment. Once fishing activity has been established the **vmstools** package then  
264 allows the analyst to explore the spatial and temporal complexity within VMS data.  
265 Based on the established link between the example tacsat and eflalo datasets, reported  
266 landings and values from the eflalo dataset were assigned to the exact positions in the  
267 tacsat dataset. Eflalo cash values were only assigned to fishing tacsat positions for which  
268 the *segmentTacsatSpeed* function was used. This function returns fishing thresholds for

269 each vessel given the gears used (see figure 2). Tacsat records with speeds between  
270 these thresholds were assumed to be fishing.

271

## 272 2.6 Spatial distribution of landings and cash value

273 Logbook declarations are made at the coarse spatial scale of the ICES statistical  
274 rectangle (1° longitude by 0.5° latitude resulting in squares of approximately ~30nm x  
275 30nm). Furthermore, the locations reported in the logbook are sometimes incorrect for a  
276 range of possible reasons (Gerritsen and Lordan, 2011). A sensible solution, then, is to  
277 exploit the connection between eflalo and tacsat to distribute landings and cash values  
278 from the logbooks at the much higher spatial (and probably more accurate) and temporal  
279 resolutions in VMS. There are, however, different aggregation levels at which these  
280 landings and cash values might be distributed among the tacsat fishing points (Bastardie  
281 et al., 2010b; Gerritsen and Lordan; Poos and Rijnsdorp, 2007). One method  
282 incorporated in the **vmstools** package (*splitAmongPings*) distinguishes three different  
283 orders, each with two or three levels. The first order is that of a full match between eflalo  
284 and tacsat using vessel IDs and trip numbers. The second order implies a match only on  
285 vessel ID, while third order means that no matching on vessel ID or trip number was  
286 possible. For first order matches, landings and cash values can be distributed among the  
287 tacsat positions, that were identified fishing, at all the various levels. These can be:  
288 landing day, landing ICES rectangle or trip number only; or a combination of these three.  
289 For second order situations, it is clear that only matches based on landing day and  
290 landing rectangle are possible. In the case of third order matches, however, distribution  
291 of landings and/or cash values can take place only between matches by landing day or  
292 landing rectangle. If no match can be found at any of the orders described, landings  
293 and/or cash values are, perforce, uniformly distributed among all tacsat pings within a  
294 year. In most occasions, however, only tacsat pings are used in which a fishing activity is  
295 assumed (see Appendix C for an overview scheme).

296

297 The allocation of cash values and landings to the example tacsat and eflalo dataset was  
298 carried out according to the following hierarchy: (1) a full match on date, ICES rectangle

299 and vessel IDs; (2) a partial match between ICES rectangle and vessel IDs; and (3) a  
300 weak match using only vessel IDs. Those eflalo records that could not be linked to any  
301 tacsat record at all are assigned first to tacsat records with similar vessel ID, following  
302 identical hierarchical levels, while those records without even similar vessel IDs are only  
303 assigned to records with matching landing date and ICES rectangle. Using this protocol  
304 we ensure that no cash-values or landings from the eflalo data are lost (see figure 3).

305

## 306 2.7 Interpolation and uncertainty

307 It can be informative to interpolate between the one or two hour interval tacsat positions  
308 to e.g. calculate area swept by mobile bottom gears or identify the origin of catches.  
309 Different interpolation techniques have been developed (Hintzen et al., 2010) of which  
310 straight line interpolation and the cubic Hermite spline method are embedded within the  
311 **vmstools** package (*interpolateTacsat*, *interpolation2Tacsat*). These methods can be used  
312 either in combination with an uncertainty estimator of possible trawling activity  
313 (*calculateCI*), or with methods for representing trawling tracks at their actual gear widths  
314 (*addWidth*).

315 From the example tacsat dataset, fishing tracks were reconstructed using the  
316 interpolation routines available in the **vmstools** package. This routine has, as yet, only  
317 been parameterized for large beam trawl fisheries. For the purpose of this example,  
318 however, we applied it here to all métiers in our test data. The interpolated track can be  
319 represented as a curved line segment (via cubic Hermite spline interpolation) or as a  
320 polygon reflecting the actual width of the gear. This enables scientists, in combination  
321 with GIS applications (e.g. *Grid2KML* can output data to Google Earth), to view the real  
322 scale of trawling impact by interactively zooming in and out (see figure 4 for a static  
323 representation). Hereafter, landings values were attributed to evenly spaced interpolated  
324 positions (*interpolation2Tacsat*).

325

## 326 2.8 Spatial resolution

327 The analyses described above are all executed without the need for any pre-defined  
328 spatial resolution, as they are conducted at the scale of the individual VMS pings. For the

329 purpose of visualizing results on maps (e.g. fishing effort per spatial unit) any size spatial  
330 grid can be defined (*vmsGridCreate*, *createGrid*) using the package. **Vmstools** allows the  
331 definition of spatial grids, given any step in either the longitudinal or latitudinal  
332 directions. Alternatively, a more restricted spatial grid definition is available where each  
333 grid cell is given a unique name following the C-square notation as developed by Rees  
334 (Rees, 2006). It should also be noted here that the spatial analyses included in **vmstools**  
335 rely heavily on the **sp** (Bivand et al., 2008) and **PBSMapping** (Schnute et al., 2008) R  
336 add-on packages.

337 The aggregated tacsat and eflalo results, as presented in figures 3c is defined on  
338 approximately square spatial grid cells with longitudinal steps of 0.1° and latitudinal  
339 steps of 0.05°. A ten times more detailed spatial grid is defined for the results in figure  
340 5b,c.

341

## 342 2.9 Indicators

343 Under the Ecosystem Approach to Fisheries Management, the use of indicators to  
344 describe ecosystem status or health has gained importance over the past years and  
345 several of those suggested can only be calculated by combining information from logbook  
346 and VMS data. The EU Data Collection Framework (EC, 2008) has identified three  
347 indicators that describe the spatial extent and impact of fishing activity: "Distribution of  
348 fishing activities", "Aggregation of fishing activities", and "Areas not impacted by mobile  
349 bottom gears" all of which can be estimated using the function *indicator* in the **vmstools**  
350 package. It should be remembered that the choice of the spatial and temporal resolution  
351 (month, quarter, year) is of great importance when calculating these indicators.

352 The DCF indicator 5, distribution of fishing activities, is calculated based on the interval  
353 rate between pings in the example tacsat dataset. The minimal number of hours of  
354 fishing activity to be included in the calculation, was set to 0, while the spatial grid  
355 defined had cell dimensions of 0.1° longitude to 0.05° latitude. To calculate the surface,  
356 the 'Trapezoid' option was used over a more accurate but slower UTM (Universal  
357 Transverse Mercator) projection option.

358

## 359 2.10 Visualisation

360 "A picture is worth a thousand words" is a particularly apt expression in the case of  
361 combined VMS and logbook analyses. The presentation of these data on maps which  
362 include geographic features such as coastlines and depth contours is extremely useful,  
363 and the **vmstools** package contains a rich suite of programs for facilitating such  
364 visualisation. Mapping routines have been developed to simplify the visualisation of  
365 tacsat and eflalo datasets, supported by existing routines from the **sp** package (e.g.  
366 *mapGrid*, *vmsGridCreate*, *plotTools*). These maps can be examined with the standard R  
367 plotting functions, while exports to other common spatial data formats are also possible.  
368 *Grid2KML*, for example, enables data to be examined in Google Earth, allowing users to  
369 interactively zoom in/out. Functionality is also available for creating animated GIFs  
370 (Graphics Interchange Format) directly from within R, via sequences of plots  
371 (*landingsMaps2GIFanim* using the **animation** R add-on package).

372

## 373 2.11 Regional databases

374 In many situations it is important to be able to assess the impact of fisheries at pan-  
375 European scales, i.e. by combining data from many EU member states. The  
376 confidentiality of these commercial data (both VMS and logbook), however, means that  
377 raw data (both VMS and logbook) will not be distributed freely among EU member states  
378 in the short term, making integrated analyses impossible. The only realistic solution,  
379 therefore, is to combine data from different countries in an aggregated format. In this  
380 context, the data warehouse for regional databases FishFrame (ICES, 2009) can now  
381 accommodate aggregated tacsat and eflalo data (see Appendix D for details). The  
382 program *pings2Fishframe* converts combined eflalo and tacsat data into the format used  
383 by FishFrame. Within FishFrame, reporting tools then allow users to display and extract  
384 any subset or combination of data required. As a proof of concept, FishFrame has been  
385 populated with subsets of Dutch and Danish data using tacsat, eflalo and the **vmstools**  
386 package.

387 FishFrame can be accessed over the web (<http://www.fishframe.org>). An extraction has  
388 been made using the 'Data Output' option, while selecting only Danish landing weights in

389 the year 2010 from the available landings VMS report. A C-square spatial grid was used  
390 to plot the total weight in the VMS data. The final figure was exported from the  
391 FishFrame web-interface and included in this paper which is shown in figure 6.

#### 392 2.11 vmstools availability and testing

393 In Table 2, a short overview of software availability and system requirements is given. To  
394 ensure the quality of the **vmstools** software, different methods have already been  
395 extensively tested and published in peer-reviewed journals (i.e. Bastardie et al., 2010b;  
396 Gerritsen and Lordan, 2011; Hintzen et al., 2010). Further testing is promoted through  
397 the use of the embedded example tacsat and eflalo which enables reliable and repeatable  
398 testing. Thereby, a manual page is written for each function available, which can be  
399 accessed at the command line in R or as a printable digital document (also available at  
400 <http://code.google.com/p/vmstools/downloads/list>), which includes an example of the  
401 function tested to operate properly when 'compiling' the R package. At the time of writing  
402 78 functions are available in **vmstools**.

403

### 404 **3. Results**

405 The general methods as described above have been applied to the example tacsat and  
406 eflalo dataset. As the datasets only comprise a subset of total activity of Dutch vessels  
407 over a two year period, no conclusions or remarks are drawn on the basis of actual  
408 patterns observed. The tables and figures are for illustration purposes only to present the  
409 capabilities of the software.

410

#### 411 3.1 Cleaning the data

412 Table 1 lists the number of records in the tacsat and eflalo datasets that were removed  
413 or flagged as they were regarded to be incorrect. In total 15% of the total tacsat records  
414 and 3% of the eflalo records were flagged or removed. Contrary to our findings in the  
415 example, in general tacsat datasets contain many duplicate records where either the GPS  
416 transponder malfunctioned or the storage of records was processed incorrectly. Due to  
417 pre-analyses to construct the example tacsat dataset, most of these have been removed.

418



### 419 3.2 Métier identification

420 For the example eflalo data, the HAC method (i) selects 30 of the 78 initial species, the  
421 second method (ii) nine, and the third (iii) 31 species. The combination of these sets  
422 defines the species retained. Within this exercise, 32 species which represent 98.9% of  
423 the total catch are retained for further analyses (*extractTableMainSpecies*). The PCA  
424 analyses of the "70% of the initial inertia" criteria indicates that 21 axes are needed to  
425 retain this threshold. Clusters of similar logbook events characterized by species  
426 assemblage are processed using the *getMetierClusters*, where the clustering method  
427 CLARA has been used to identify the métiers of the example eflalo dataset. This method  
428 proposes a classification in four clusters characterized by one or more species. The  
429 clusters count respectively 4040, 69, 244 and 186 logbook events. Cluster one  
430 corresponds to a flatfish métier, listing sole and plaice as main target species. Cluster  
431 two lists eel and lobster while cluster three is characterized by mackerel and Horse  
432 mackerel. The fourth cluster is characterized by sea bass and mullets (see figure 7).  
433 Hereafter, the eflalo dataset is complemented with a column indicating the métier  
434 identification per logbook event.

435

### 436 3.3 Linking eflalo to tacsat data

437 In total 96% of all tacsat records could be linked to an eflalo record. The results of this  
438 process are shown in figure 3a which represents landings from eflalo data by ICES  
439 statistical rectangle. Landings are then assigned to tacsat positions and the output  
440 summed over the same grid (ICES statistical rectangles, figure 3b); while in the last map  
441 (figure 3c) the same data is aggregated over a finer grid. Clearly, while the first two plots  
442 are rather similar a totally different understanding of landings by this fleet is gained  
443 when examining the data at the finer scales.

444

### 445 3.4 Fishing activity

446 Fishing activity has been determined for all vessels, however, figure 2 only shows the  
447 results for vessel ID 298138. The analyses, based on speeds, computed from the  
448 Euclidian distance and time difference between VMS pings, rather than instantaneous

449 speed, results in a fishing speed upper and lower boundary of respectively 3.29 and 5.31  
450 knots per hour. Figure 2a represents the cumulative and pre-smoothed distribution of  
451 speeds employed by the vessel, while figure 2b represents the instantaneous speed  
452 distribution where figure 2c represents the calculated speed distribution.

453

### 454 3.5 Reconstruction of fishing trips with higher ping-rates

455 In total 11668 fishing tracks were reconstructed from the tacsat dataset, noting that  
456 reconstruction is set to take place only when fishing positions are at most 2 hours apart.  
457 In our test data this amounts to 46% of all tacsat records classified as fishing.  
458 Figure 5a presents the fine scale distribution of landings of herring, plaice and sole in the  
459 North sea, identical to figure 3c. The difference between panel (5b) and (5c) shows how  
460 the information-density and perception of spatial impact can differ when using different  
461 track reconstruction techniques. Figure 4a shows the tacsat positions of vessel 157  
462 employing a beamtrawl gear. Figure 4b shows the same vessel, however, now after  
463 reconstruction of fishing trips. Adding a width to the same interpolation (*addWidth*)  
464 results in figure 4c and 4d representing the actual width of the gear.

465

### 466 3.6 Indicators

467 A graphical representation of distribution of fishing activity, calculated based on the  
468 example tacsat and eflalo dataset, by month, as follows from Table 3, is given in figure 8.  
469 The activity is defined as the number of hours of fishing activity in each grid cell.

470

### 471 3.7 Fishframe

472 FishFrame is only populated with real data. Hence, the example given in figure 6  
473 represents Danish landings data in 2010 only of species cod. The pre-processing to  
474 populate FishFrame with these data was executed using the **vmstools** package.

475

## 476 **4. Discussion**

477

478 Vessel Monitoring by Satellite system data are potentially valuable for quantifying the  
479 activity and impact of fishing on the marine environment. They can be used to inform  
480 spatial planning, to address conservation and biodiversity management and to monitor  
481 seafloor integrity and bottom impact of fishing (CEC, 2007). This study demonstrates  
482 how VMS data and logbook analyses can be performed in an efficient and standardized  
483 manner using the **vmstools** package in conjunction with R. Although the literature  
484 provides examples of similar procedures, analyses and visualizations of VMS and logbook  
485 data (Bastardie et al., 2010b; Fock, 2008; Gerritsen and Lordan, 2011; Hintzen et al.,  
486 2010; ICES, 2010; Lee et al., 2010; Pedersen et al., 2009), no framework had yet been  
487 presented that enables such analyses to take place in a consistent manner via  
488 transparent open-access tools. Similarly, while many previous studies have applied  
489 clustering methods to define métiers at the national level (Marchal, 2008; Ulrich et al.,  
490 2009), no generic framework for comparing these methods and suggesting a complete,  
491 objective and operational workflow for the analyses of landings profiles at the supra-  
492 national (regional) level had yet been implemented. The strength of the **vmstools**  
493 package is that it has built upon the existing, but isolated tools, to provide such a unified  
494 framework. Hence, the **vmstools** library is qualified to respond to the demand for  
495 standard and transparent procedures that can be done by different analysts, and is  
496 already recommended as the basis for future work on VMS and logbook analyses by ICES  
497 (ICES, 2011).

498

499 Here, R was chosen as the development environment for a variety of reasons. In Europe,  
500 many fisheries institutes are familiar with its operation, secondly it is freely available, and  
501 thirdly it has an ever-growing list of additional libraries (Thyer et al., 2011) which can be  
502 utilized within the VMS and logbook analyses. It is often criticised for being  
503 computationally slow in comparison with other systems especially when dealing with  
504 large datasets such as VMS data. In constructing generic routines, however, we ensured  
505 that the algorithms are as efficient as possible and indeed most jobs can be achieved  
506 quickly and hence we believe that the advantages of R outweigh its disadvantages. For  
507 example, there is no need to switch between different programs to complete a back-to-

508 back analyses, or use complicated interfaces that allow different software programs to  
509 communicate with each other, in contrast to other geospatial tools where a suite of  
510 programs is used to improve speed performance of the analyses (Roberts et al., 2010;  
511 Cagnacci and Urbano, 2008). Further these are often prone to failures when updates are  
512 released (Roberts et al., 2010). Moreover, documentation of the generic functions is  
513 straightforward, as is the creation of help files and the provision of example scripts. The  
514 mapping facilities in R are also very powerful. However, although many of the generic  
515 functions take the spherical shape of the earth into account (utilizing additional libraries),  
516 it is to be noted that no true GIS environment is mimicked, and this may, in some  
517 situations result in small differences, as in most calculations the earth is assumed to be  
518 described by a perfect sphere. Especially when decreasing the spatial resolution, the  
519 differences become significantly different.

520

521 Key to the start of **vmstools** was an agreement on common data formats. The data  
522 formats (eflalo, tacsat) were chosen on the basis that many European fisheries institutes  
523 had previous expertise in using them for exchanging scientific programs and data. Hence,  
524 it was viewed as a practical solution to implement these formats as standard and on  
525 which all functions developed rely. The use of these standardized data format can and  
526 has facilitated the development of the software greatly (Kranstauber et al., 2011) and  
527 also ensures users of a degree of quality assurance marked as important by ICES (ICES,  
528 2010). Moreover, the standardized input files can be processed with **vmstools** and then  
529 output files can be created which can be uploaded into FishFrame. Once the data are in  
530 FishFrame, data from different countries all over the world can then be integrated to  
531 produce regional maps. The advantage of this approach is that each country always  
532 keeps its raw VMS and logbook data locally, while also being able to exploit the more  
533 comprehensive coverage available in the regional database of aggregated data. These  
534 aggregated VMS and logbook data can then be linked to other fishery-related data types,  
535 e.g. survey information, biological sampling or discard information.

536

537 Providing software with built-in functionality to efficiently process and analyse VMS and  
538 logbook data also has its downside. Scientists, who are not aware of the implications of  
539 the raw VMS and logbook data, can potentially misinterpret the results provided via one  
540 of the algorithms. The ability, for example, to distribute landings and the associated  
541 cash-values from logbooks among VMS positions is still controversial; although such  
542 analyses are already appearing in the literature (Gerritsen and Lordan, 2011). The  
543 procedure can give the impression that landings were indeed taken from these very  
544 localised positions, and this may be misleading since it is merely an interpretation of  
545 independent sources, and thus it depends much on our ability to merge the VMS data  
546 successfully with the logbooks and to estimate real fishing activity.

547  
548 Developing software in science should facilitate the scientific process to either  
549 standardize certain processes or make them more efficient. The framework described  
550 here enables scientists to easily combine detailed information from e.g. companies laying  
551 cables, oil and gas exploration, or wind farms with fishing activity information which can  
552 be used to inform, at high spatial scale, models that study spatially explicit fisheries  
553 behaviour and their impact on marine ecosystems (Bastardie et al., 2010a). The spatial  
554 detail comprised by the combined information might be used to link environmental  
555 drivers to fisheries behaviour, or even to relate sea mammal tracking data with fishing  
556 data for e.g. spatial overlap studies to determine human-mammal competition  
557 (Matthiopoulos et al., 2008) or sea bird movement to fishing activity (Copello and  
558 Quintana, 2009) in and outside closed areas (Trebilco et al., 2008). Especially when  
559 dealing with spatial management, the information obtained from linked VMS and logbook  
560 data plays an important role (Dinmore, 2003; Fock, 2008; Murawski et al., 2005;  
561 Pedersen et al., 2009; Stelzenmuller et al., 2008) where spatial activity information is  
562 used to facilitate the process in designing marine protected areas or inform stakeholders  
563 on fishing impact. Similarly, grouping fishing activities into limited numbers of métiers at  
564 the regional scale is the first necessary step towards integrated fleet-based fisheries  
565 management, allowing moving beyond the current single-stock management schemes  
566 when fisheries are mixed (Ulrich et al., 2009). New insights in these fields, supported by

567 technical advances could easily be incorporated exposing scientific breakthroughs to a  
568 much larger public and provide a transition from output to outcome oriented science  
569 where applications may change how spatial management is organized (Matthews et al.,  
570 2011). Hence, the authors hope that the **vmstools** framework will allow the user to  
571 focus on (a better understanding of) the impact of human activities on the marine  
572 ecosystem where the knowledge gained, not necessarily the software, plays a central role  
573 (Matthews et al., 2011).

574

## 575 **5. Supplementary material**

576 Supplementary material is available at the Fisheries Research online version of the paper  
577 and includes a detailed description of the eflalo and tacsat data formats, the R-script  
578 used to perform all analyses as shown in this paper, a graphical overview of the linking  
579 process between eflalo and tacsat and a detailed description of the FishFrame format.

580

## 581 **Acknowledgements**

582 This study was funded by the Commission of the European Communities under the call  
583 for tender MARE/2008/10; Lot 2 - Development of tools for logbook and VMS data  
584 analysis. We thank all other members, Fabrizio Manco, Hans Gerritsen, Sebastien  
585 Demanèche, Stéphanie Mahévas, Martial Laurans, Neil Campbell, J. Rasmus Nielsen,  
586 Stuart A. Reeves, Colm Lordan and Ryszard Grzebielec for their efforts within the project.

587

## 588 **References**

- 589 Bastardie, F., Nielsen, J.R. Andersen, B.S., and Eigaard, O.R. 2010a. Effects of fishing  
590 effort allocation scenarios on energy efficiency and profitability: An individual-  
591 based model applied to Danish fisheries. Fisheries Research 106, 501-516.
- 592 Bastardie, F., Nielsen, J.R., Ulrich, C., Egekvist, J., Degel, H., 2010b. Detailed mapping of  
593 fishing effort and landings by coupling fishing logbooks with satellite-recorded  
594 vessel geo-location. Fisheries Research 106, 41-53.
- 595 Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V., 2008. Applied Spatial Data Analysis with

596 R. Springer.

597 Cagnacci, F., Urbano, F., 2008. Managing wildlife: A spatial information system for GPS  
598 collars data. *Environmental Modelling & Software* 23, 957-959.

599 CEC, 2007. European Parliament legislative resolution of 11 December 2007 on the  
600 Council common position for adopting a directive of the European Parliament and  
601 of the Council establishing a Framework for Community Action in the field of  
602 Marine Environmental Policy.

603 Christensen, V., Walters, C.J., 2004. Ecopath with Ecosim: methods, capabilities and  
604 limitations. *Ecological Modelling* 172, 109-139.

605 Copello, S., Quintana, F., 2009. Spatio-temporal overlap between the at-sea distribution  
606 of Southern Giant Petrels and fisheries at the Patagonian Shelf. *Polar Biology* 32,  
607 1211-1220.

608 Dinmore, T.A., 2003. Impact of a large-scale area closure on patterns of fishing  
609 disturbance and the consequences for benthic communities. *ICES Journal of*  
610 *Marine Science* 60, 371-380.

611 Deporte, N., Bastardie, F., Demanèche, S., Mahévas, S., Ulrich, C., submitted,  
612 Comparative investigation of statistical methods for identifying fishing activities  
613 from logbook data: Example of application to the international otter trawl  
614 fisheries in the North Sea.

615 Douvere, F., Ehler, C.N., 2009. New perspectives on sea use management: Initial  
616 findings from European experience with marine spatial planning. *Journal of*  
617 *Environmental Management* 90, 77-88.

618 Eastwood, P., 2007. Human activities in UK offshore waters: an assessment of direct,  
619 physical pressure on the seabed. *ICES Journal of Marine Science* 64, 453-463.

620 EC, 1993. Council Regulation (EC) No. 2847/93 of 12 October 1993 establishing a control  
621 system applicable to the common fisheries policy L261, 1-16.

622 EC, 2002. Council Regulation (EC) No. 2371/2002 of 20 December 2002 on the  
623 conservation and sustainable exploitation of fisheries resources under the  
624 Common Fisheries Policy. *Official Journal of the European Communities* L358, 59-  
625 80.

626 EC, 2008. Commission Decision of 6 November 2008 adopting a multiannual Community  
 627 programme pursuant to Council Regulation (EC) No 199/2008 establishing a  
 628 Community framework for the collection, management and use of data in the  
 629 fisheries sector and support for scientific advice regarding the common fisheries  
 630 policy (2008/949/EC) L 346, 37–88.

631 Fock, H.O., 2008. Fisheries in the context of marine spatial planning: Defining principal  
 632 areas for fisheries in the German EEZ. *Marine Policy* 32, 728-739.

633 Gerritsen, H., Lordan, C., 2011. Integrating vessel monitoring systems (VMS) data with  
 634 daily catch data from logbooks to explore the spatial distribution of catch and  
 635 effort at high resolution. *ICES Journal of Marine Science* 68, 245-252.

636 Hartigan, J.A., 1975. *Clustering Algorithms*. John Wiley & Sons, Inc., New York.

637 Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. *Applied Statistics* 28,  
 638 100-108.

639 Hintzen, N.T., Piet, G.J., Brunel, T., 2010. Improved estimation of trawling tracks using  
 640 cubic Hermite spline interpolation of position registration data. *Fisheries Research*  
 641 101, 108-115.

642 ICES, 2003. Report of the Study Group on the development of fishery-based forecasts  
 643 (SGDFF). ICES CM 2003/ACFM: 08, Copenhagen, p. 37.

644 ICES, 2009. Definition of Standard Data-Exchange Format for Sampling, Landings, and  
 645 Effort Data from Commercial Fisheries. ICES Cooperative Research Report,  
 646 Copenhagen, p. 44.

647 ICES, 2010. Report of the Study Group on VMS data, its storage, access and tools for  
 648 analysis (SGVMS). ICES CM 2010/SSGSUE:12, Copenhagen, p. 46.

649 ICES, 2011. Report on the Workshop on the utility of commercial CPUE and VMS data in  
 650 assessment (WKCPUEEFFORT). Copenhagen, p. 67.

651 Kaiser, M.J., Clarke, K.R., Hinz, H., Austen, M.C.V., Somerfield, P.J., Karakassis, I., 2006.  
 652 Global analysis of response and recovery of benthic biota to fishing. *Marine*  
 653 *Ecology Progress Series* 311, 1-14.

654 Kaufman, L., Rousseeuw, P.J., 2008. *Introduction, Finding Groups in Data*. John Wiley &  
 655 Sons, Inc., New York, pp. 1-67.



656 Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., D., G., Hillary, R., Jardim, E.,  
657 Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., Scott, R.D., 2007. FLR: an open-  
658 source framework for the evaluation and development of management strategies.  
659 ICES Journal of Marine Science 64, 640-646.

660 Kraak, S.B.M., Buisman, F.C., Dickey-Collas, M., Poos, J.J., Pastoors, M.A., Smit, J.G.P.,  
661 van Oostenbrugge, J.A.E., Daan, N., 2008. The effect of management choices on  
662 the sustainability and economic performance of a mixed fishery: a simulation  
663 study. ICES Journal of Marine Science 65, 697-712.

664 Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M., Kays, R.,  
665 2011. The Movebank data model for animal tracking. Environmental Modelling &  
666 Software 26, 834-835.

667 Lee, J., South, A.B., Jennings, S., 2010. Developing reliable, repeatable, and accessible  
668 methods to provide high-resolution estimates of fishing-effort distributions from  
669 vessel monitoring system (VMS) data. ICES Journal of Marine Science 67, 1260-  
670 1271.

671 Long, R.J., Curran, P.A., 2000. Enforcing the common fisheries policy. Oxford: Fishing  
672 News Books.

673 Marchal, P., 2008. A comparative analysis of métiers and catch profiles for some French  
674 demersal and pelagic fleets. ICES Journal of Marine Science 65, 674-686.

675 Matthews, K.B., Rivington, M., Blackstock, K., McCrum, G., Buchan, K., Miller, D.G.,  
676 2011. Raising the bar? - The challenges of evaluating the outcomes of  
677 environmental modelling and software. Environmental Modelling & Software 26,  
678 247-257.

679 Matthiopoulos, J., Smout, S., Winship, A.J., Thompson, D., Boyd, I.L., Harwood, J., 2008.  
680 Getting beneath the surface of marine mammal – fisheries competition. Mammal  
681 Review 38, 167-188.

682 Mills, C.M., Townsend, S.E., Jennings, S., Eastwood, P.D., Houghton, C.A., 2007.  
683 Estimating high resolution trawl fishing effort from satellite-based vessel  
684 monitoring system data. ICES Journal of Marine Science 64, 248-255.

685 Murawski, S.A., Wigley, S.E., Fogarty, M.J., Rago, P.J., Mountain, D.G., 2005. Effort

686 distribution and catch patterns adjacent to temperate MPAs. *ICES Journal of*  
687 *Marine Science* 62, 1150-1167.

688 Pastoors, M.A., Poos, J.J., Kraak, S.B.M., Machiels, M.A.M., 2007. Validating  
689 management simulation models and implications for communicating results to  
690 stakeholders. *ICES Journal of Marine Science* 64, 818-824.

691 Pedersen, S.A., Fock, H.O., Sell, A.F., 2009. Mapping fisheries in the German exclusive  
692 economic zone with special reference to offshore Natura 2000 sites. *Marine Policy*  
693 33, 571-590.

694 Piet, G.J., Quirijns, F.J., 2009. The importance of scale for fishing impact estimations.  
695 *Canadian Journal of Fisheries and Aquatic Sciences* 66, 829-835.

696 Piet, G.J., Quirijns, F.J., Robinson, L., Greenstreet, S.P.R., 2007. Potential pressure  
697 indicators for fishing, and their data requirements. *ICES Journal of Marine Science*  
698 64, 110-121.

699 Poos, J.-J., Rijnsdorp, A.D., 2007. The dynamics of small-scale patchiness of plaice and  
700 sole as reflected in the catch rates of the Dutch beam trawl fleet and its  
701 implications for the fleet dynamics. *Journal of Sea Research* 58, 100-112.

702 R Development Core Team, 2008. R: A language and environment for statistical  
703 computing. R Foundation for Statistical Computing, Vienna, Austria.

704 Rees, T., 2006. "C-squares", a new spatial indexing system and its applicability to the  
705 description of oceanographic datasets. *Oceanography* 16, 11-19.

706 Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and Opportunities of  
707 Open Data in Ecology. *Science* 331, 703-705.

708 Rijnsdorp, A.D., Buys, A.M., Storbeck, F., Visser, E.G., 1998. Micro-scale distribution of  
709 beam trawl effort in the southern North Sea between 1993 and 1996 in relation to  
710 the trawling frequency of the sea bed and the impact on benthic organisms. *ICES*  
711 *Journal of Marine Science* 55, 403-419.

712 Roberts, J.J., Best, B.D., Dunn, D.C., Trembl, E.A., Halpin, P.N., 2010. Marine Geospatial  
713 Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS,  
714 Python, R, MATLAB, and C++. *Environmental Modelling & Software* 25, 1197-  
715 1207.

716 Sainsbury, K.J., Punt, A.E., Smith, A.D.M., 2000. Design of operational management  
717 strategies for achieving fishery ecosystem objectives. ICES Journal of Marine  
718 Science 57, 731-741.

719 Schnute, J.T., Boers, N.M., Haigh, R., Couture-Beil, A., 2008. PBS Mapping 2.57: User's  
720 Guide - revised from *Canadian Technical Report of Fisheries and Aquatic Sciences*  
721 2549. vi +, p. 118.

722 Stelzenmuller, V., Rogers, S.I., Mills, C.M., 2008. Spatio-temporal patterns of fishing  
723 pressure on UK marine landscapes, and their implications for spatial planning and  
724 management. ICES Journal of Marine Science 65, 1081-1091.

725 Thyer, M., Leonard, M., Kavetski, D., Need, S., Renard, B., 2011. The open source  
726 RFortran library for accessing R from Fortran, with applications in environmental  
727 modelling. Environmental Modelling & Software 26, 219-234.

728 Trebilco, R., Gales, R., Baker, G.B., Terauds, A., Sumner, M.D., 2008. At sea movement  
729 of Macquarie Island giant petrels: Relationships with marine protected areas and  
730 Regional Fisheries Management Organisations. Biological Conservation 141, 2942-  
731 2958.

732 Ulrich, C., Wilson, D.C., Nielsen, J.R., Reeves, S.A., 2009. Potentials and challenges in  
733 fleet- and métier- based approaches for fisheries management in the CFP. ICES  
734 CM 2009/R: 06, p. 23.

735 Vermard, Y., Rivot, E., Mahévas, S., Marchal, P., Gascuel, D., 2010. Identifying fishing  
736 trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden  
737 Markov Models. Ecological Modelling 221, 1757-1769.

738 Walker, E., Bez, N., 2010. A pioneer validation of a state-space model of vessel  
739 trajectories (VMS) with observers' data. Ecological Modelling 221, 2008-2017.

740

## 741 **Table and Figure captions**

742 **Table 1:** Overview of processing and filtering results of the example tacsat and eflalo data files. Original  
743 number of rows in each dataset is given, while percentages removed / flagged are calculated based on the  
744 original number of entry rows.

745

746 **Table 2:** Overview of software availability and system requirements

747  
748 **Table 3:** Results by month of DCF indicator 5 calculation on three different spatial scales. Spatial scales are  
749 given in degrees longitude and degrees latitude respectively. Values denote the total surface (in km<sup>2</sup>) with  
750 fishing activity by month as calculated as the sum of the grid cells surface with fishing activity.

751  
752 **Figure 1:** Panels show part of the North Sea where in panel (a) red dots represent all vessel positions as listed  
753 in the tacsat file. In panel (b), dark blue dots represent all points that are flagged 'in harbour' which have been  
754 selected based on a 3km radius from a chosen central point in each European harbour. In panel (c) all vessel  
755 locations that are located on land are represented by dark orange dots. The analyses are based on the example  
756 tacsat and eflalo datasets available in the **vmstools** package. (NOT FOR PUBLICATION: color on web, black-  
757 and-white in print)

758  
759 **Figure 2:** Segmented regression analyses of vessel (ID = "298138", gear = OTM) where panel (a) represents  
760 the cumulative distribution of calculated speeds. The blue vertical dotted lines represent the speed thresholds  
761 (at 3.29 and 5.31 knots per hour) as identified by the segmented regression in which the vessel is assumed to  
762 be fishing. Panel (b) shows the instantaneous speed distribution as provided by the VMS data while panel (c)  
763 shows the speed distribution calculated given the Euclidian distance and time interval between successive VMS  
764 pings. The analyses are based on the example tacsat and eflalo datasets available in the **vmstools** package.  
765 (NOT FOR PUBLICATION: color on web, black-and-white in print)

766  
767 **Figure 3:** Panels show aggregated and standardized (for confidentiality purposes) landings of herring, plaice  
768 and sole in (part of) the North Sea where in panel (a) landings from eflalo by ICES statistical rectangle (the  
769 highest spatial detail possible, 1 x 0.5 grid cell in respectively longitudinal and latitudinal degrees) are shown;  
770 Panel (b) shows landings after linking tacsat and eflalo by ICES statistical and distributing landings from eflalo  
771 according to tacsat positions. Panel (c) shows landings after linking tacsat and eflalo by 0.1 x 0.05 grid cell in  
772 respectively longitudinal and latitudinal degrees. The analyses are based on the example tacsat and eflalo  
773 datasets available in the **vmstools** package. (NOT FOR PUBLICATION: color on web, black-and-white in print)

774  
775 **Figure 4:** Panels show the distribution of a single fishing vessel in the southern North Sea (ID = "157", gear =  
776 TBB) where panel (a) represents all the VMS pings available in the test tacsat dataset in **vmstools** after  
777 filtering erroneous records. Panel (b) shows the interpolated fishing tracks where non-fishing behaviour has  
778 been filtered out. Panel (c) shows a similar interpolation, however, added with a representative gear width. As  
779 this gear width only stretches 24 meters, it is hardly visible on the large North Sea scale, hence, panel (d)  
780 represents an enlargement of a smaller area where the width of the gear is spatially shown. The analyses are  
781 based on the example tacsat and eflalo datasets available in the **vmstools** package. (NOT FOR PUBLICATION:  
782 color on web, black-and-white in print)

783

784 **Figure 5:** Panels show aggregated and standardized (for confidentiality purposes) landings of herring, plaice  
785 and sole in (part of) the North Sea where in panel (a) the standardized landings as obtained from the tacsat  
786 dataset are shown at a scale of 0.1 x 0.05 in respectively longitudinal and latitudinal degrees. Panel (b) shows  
787 an enlargement, and finer spatial scale, of the black squared area from panel (a) at a scale of 0.01 x 0.005 in  
788 respectively longitudinal and latitudinal degrees. Panel (c), also shows an enlargement at a finer spatial scale of  
789 this area, but is based on a tacsat dataset complemented with reconstructed fishing tracks where each  
790 successful track is assigned eight intermediate points. The analyses are based on the example tacsat and eflalo  
791 datasets available in the **vmstools** package. (NOT FOR PUBLICATION: color on web, black-and-white in print)

793 **Figure 6:** Landings (in kg) of Atlantic cod by the Danish fleet in the first quarter of 2010 as extracted from the  
794 online FishFrame tool. Landings are aggregated at a scale of 0.05 x 0.05 in longitudinal and latitudinal degrees  
795 (smallest C-square resolution). (NOT FOR PUBLICATION: color on web, black-and-white in print)

797 **Figure 7:** Percentage of cash value per species (FAO code) per cluster (I to IV) applying the procedure PCA &  
798 CLARA algorithm on the example eflalo data set available in the **vmstools** package. Only species labels with  
799 cash value greater than 10% within at least one cluster are displayed for clarity. Abbreviations: SOL: sole, PLE:  
800 plaice, MUL: Mulletts, MAC: mackerel, LBE: European lobster, JAX: Horse mackerel, ELE: Eel, BSS: sea bass.  
801 (NOT FOR PUBLICATION: color on web, black-and-white in print)

803 **Figure 8:** Panels show the activity of fishing by month, based on the DCF indicator 5, represented at a scale of  
804 0.1 x 0.05 in respectively longitudinal and latitudinal degrees. The aggregated indicator values can be obtained  
805 from Table 3. The analyses are based on the example tacsat and eflalo datasets available in the **vmstools**  
806 package. (NOT FOR PUBLICATION: color on web, black-and-white in print)

**Table 1:** Overview of processing and filtering results of the example tacsat and eflalo data files. Original number of rows in each dataset is given, while percentages removed / flagged are calculated based on the original number of entry rows.

Table	Processing description	% removed / flagged
Tacsat	-	97015 rows
Tacsat	Longitude, latitude and speed outside range	0.00%
Tacsat	Duplicate records removed	0%
Tacsat	Points in harbour	15.22%
Tacsat	Points on land	0.16%
Eflalo	-	4539 rows
Eflalo	Mesh size out of range	0.57%
Eflalo	Duplicate records removed	2.05%
Eflalo	Arrival date before departure date	0%

**Table 2:** Overview of software availability and system requirements

Name of software	vmstools
Developers	IMARES, DTU-Aqua, IFREMER, CEFAS
Contact	<a href="mailto:Niels.Hintzen@wur.nl">Niels.Hintzen@wur.nl</a>
Year first available	2011
Hardware recommended	2 GB RAM, 2 GHz CPU
Software required	>= R2.11.1, compiled for Unix or Windows OS
Program Language	R
Program size	~4 MB
Availability	<a href="http://code.google.com/p/vmstools/downloads/list">http://code.google.com/p/vmstools/downloads/list</a>
Cost	Free, available as package under R

**Table 3:** Results by month of DCF indicator 5 calculation on three different spatial scales. Spatial scales are given in degrees longitude and degrees latitude respectively. Values denote the total surface (in km<sup>2</sup>) with fishing activity by month as calculated as the sum of the grid cells surface with fishing activity.

Month	Spatial scale 1 – 0.5	Spatial scale 0.1 – 0.05	Spatial scale 0.01 – 0.005
1	107535	6649	100
2	88238	6866	131
3	74442	8776	301
4	127857	9201	150
5	182241	48894	1734
6	150221	17245	391
7	95719	4300	69
8	124076	9769	208
9	182898	75217	3337
10	130768	10010	136
11	112094	6355	84
12	49307	4723	86

Figure 1 Color

[Click here to download high resolution image](#)

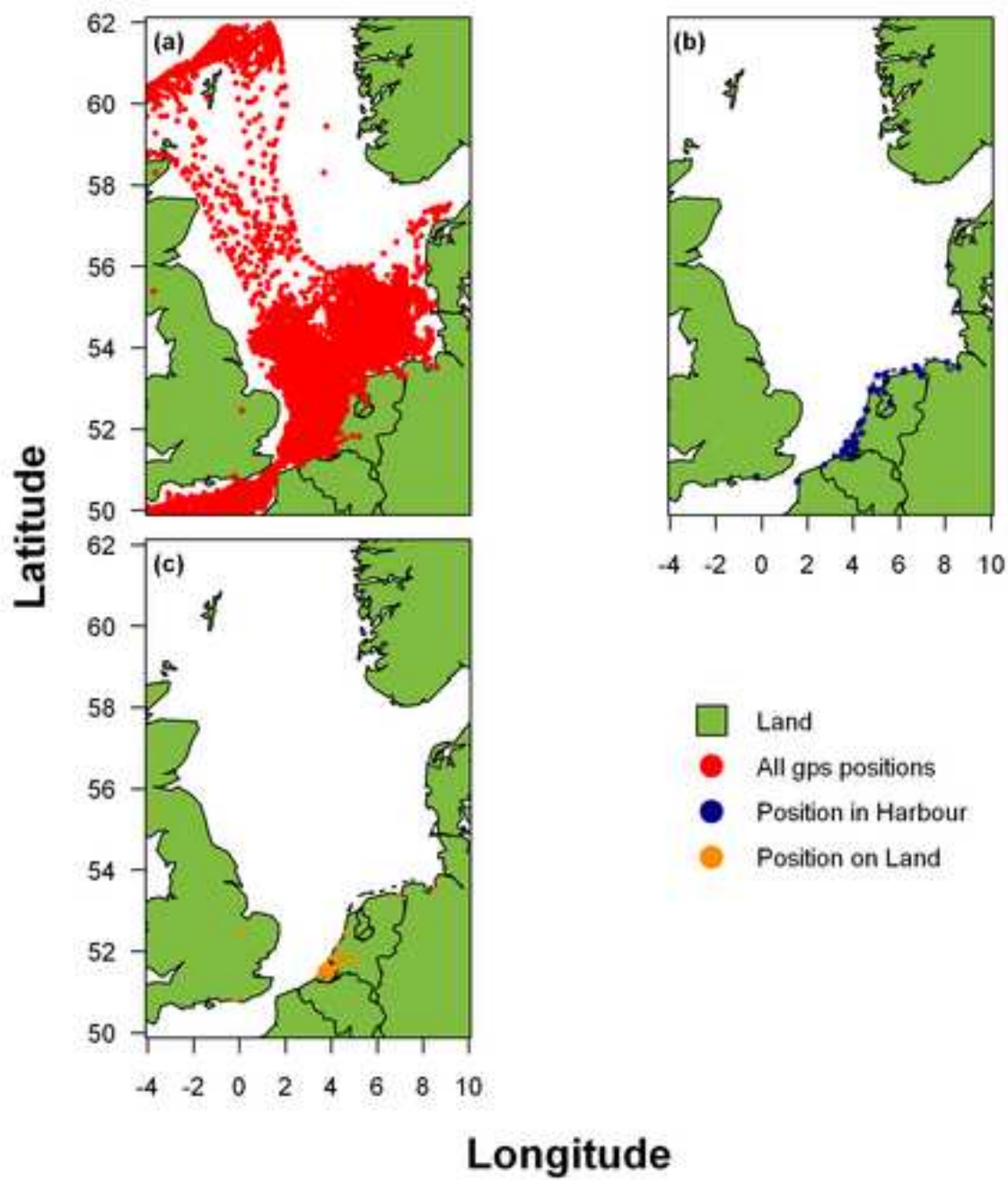


Figure 2 Color  
[Click here to download high resolution image](#)

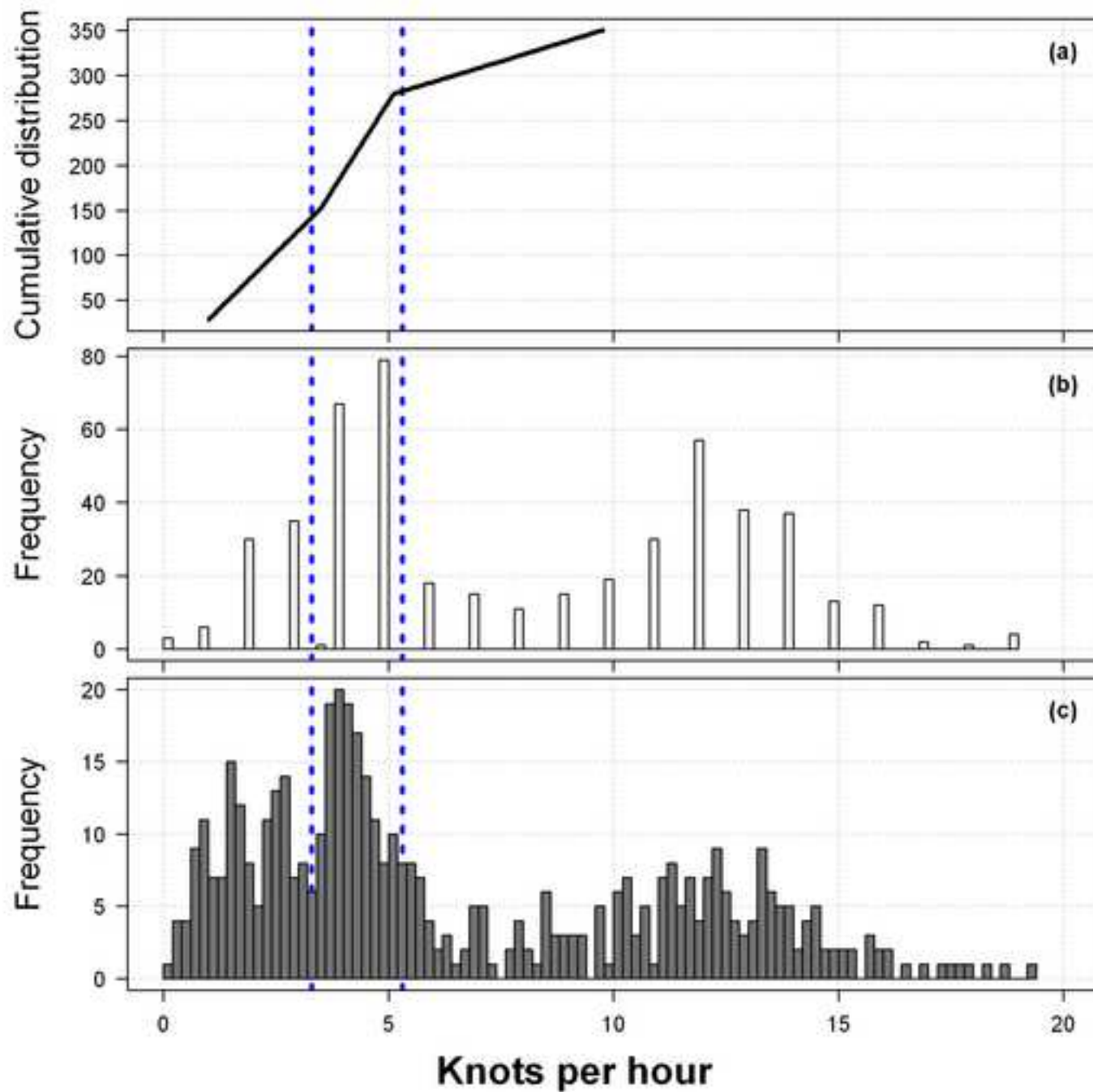




Figure 3 Color  
[Click here to download high resolution image](#)

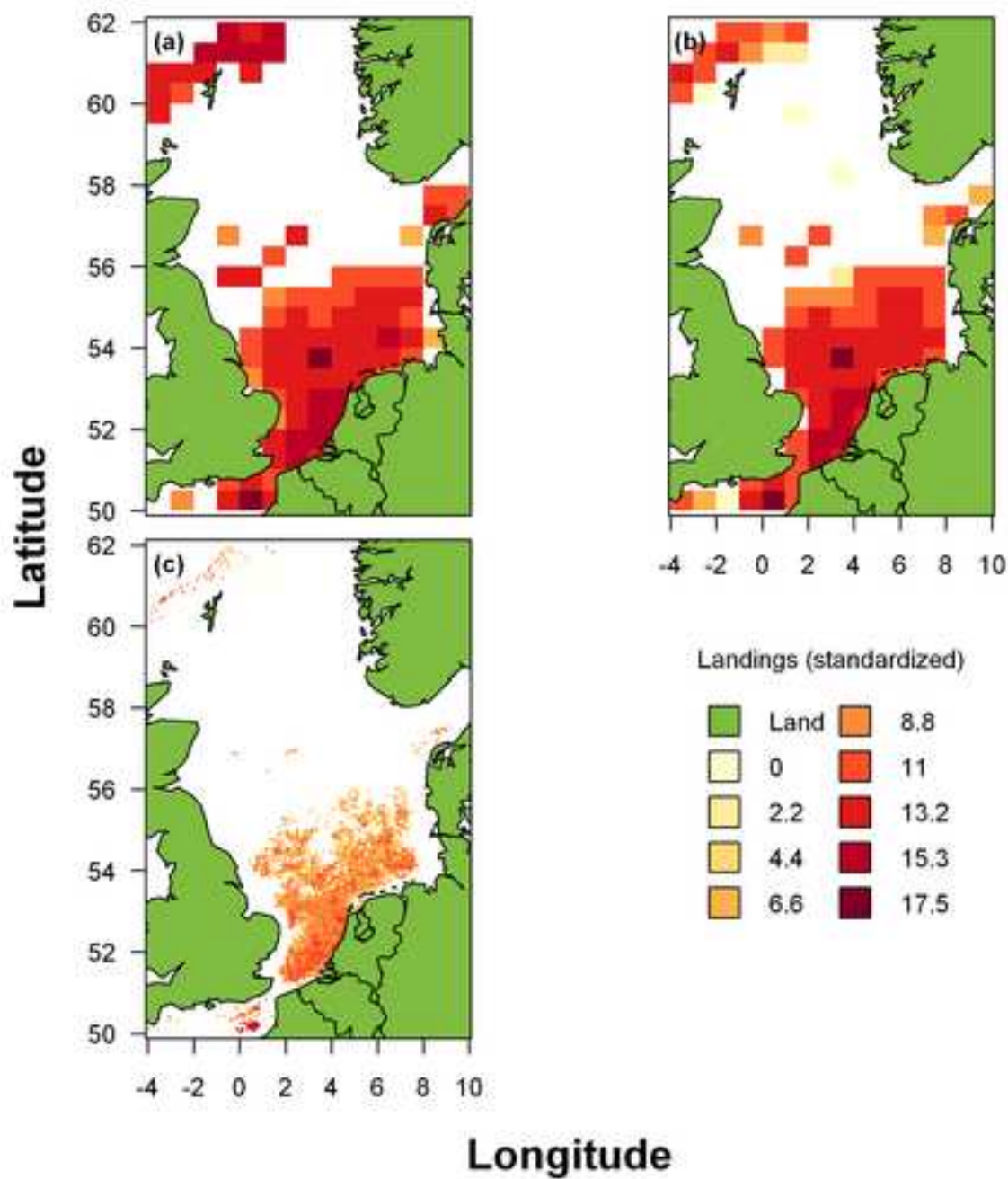


Figure 4 Color  
[Click here to download high resolution image](#)

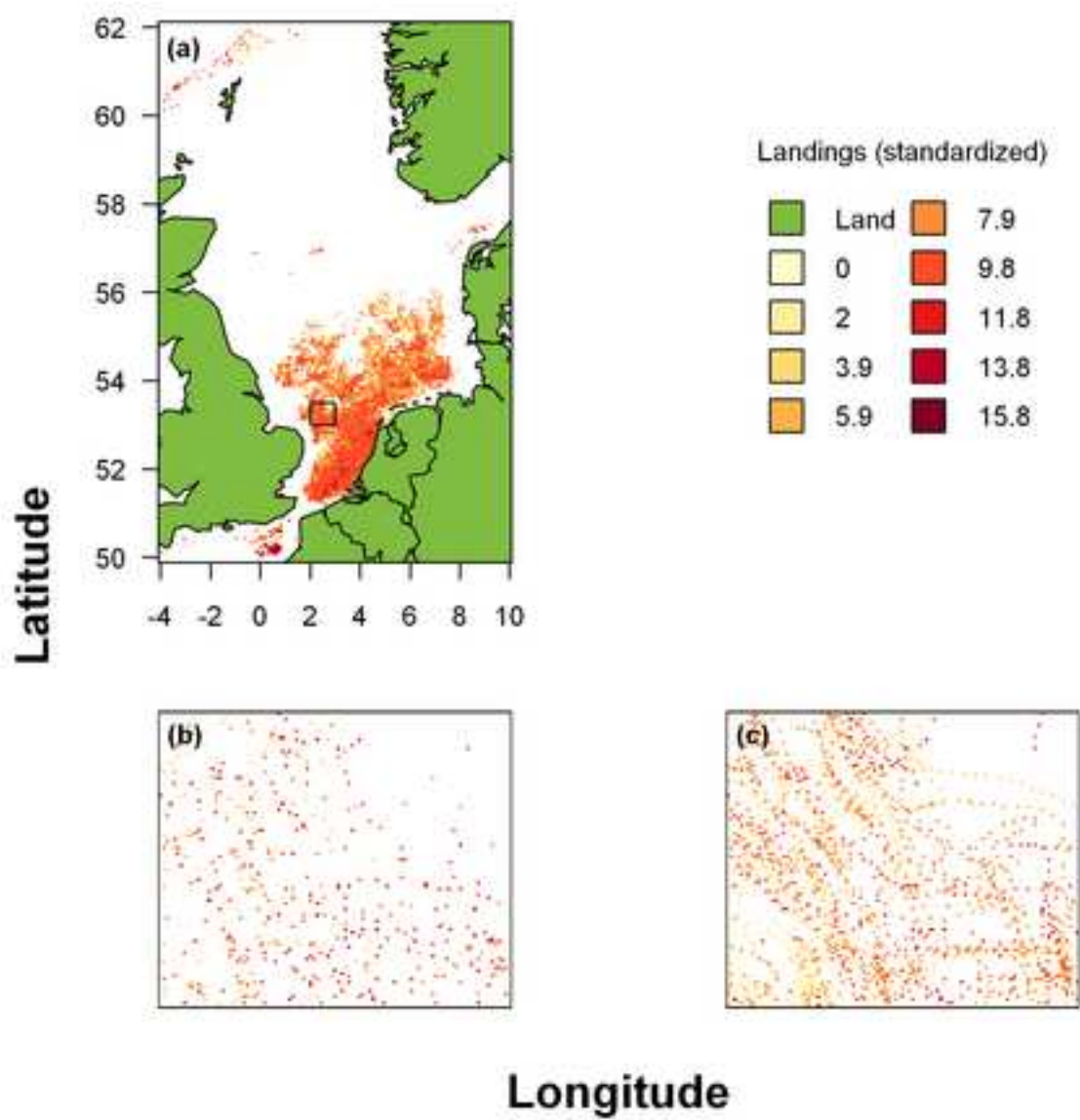


Figure 5 Color

[Click here to download high resolution image](#)

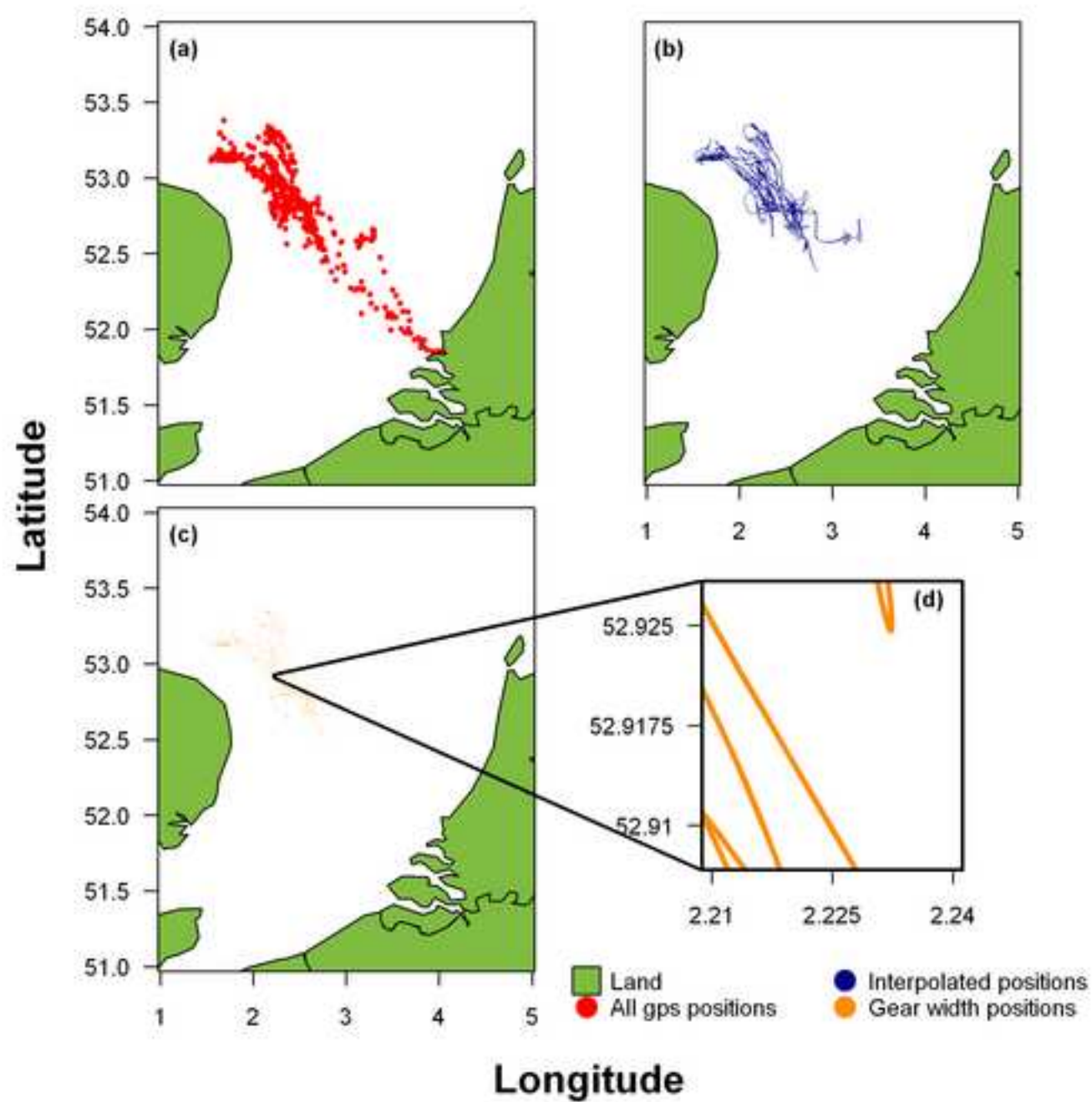


Figure 6 Color

[Click here to download high resolution image](#)

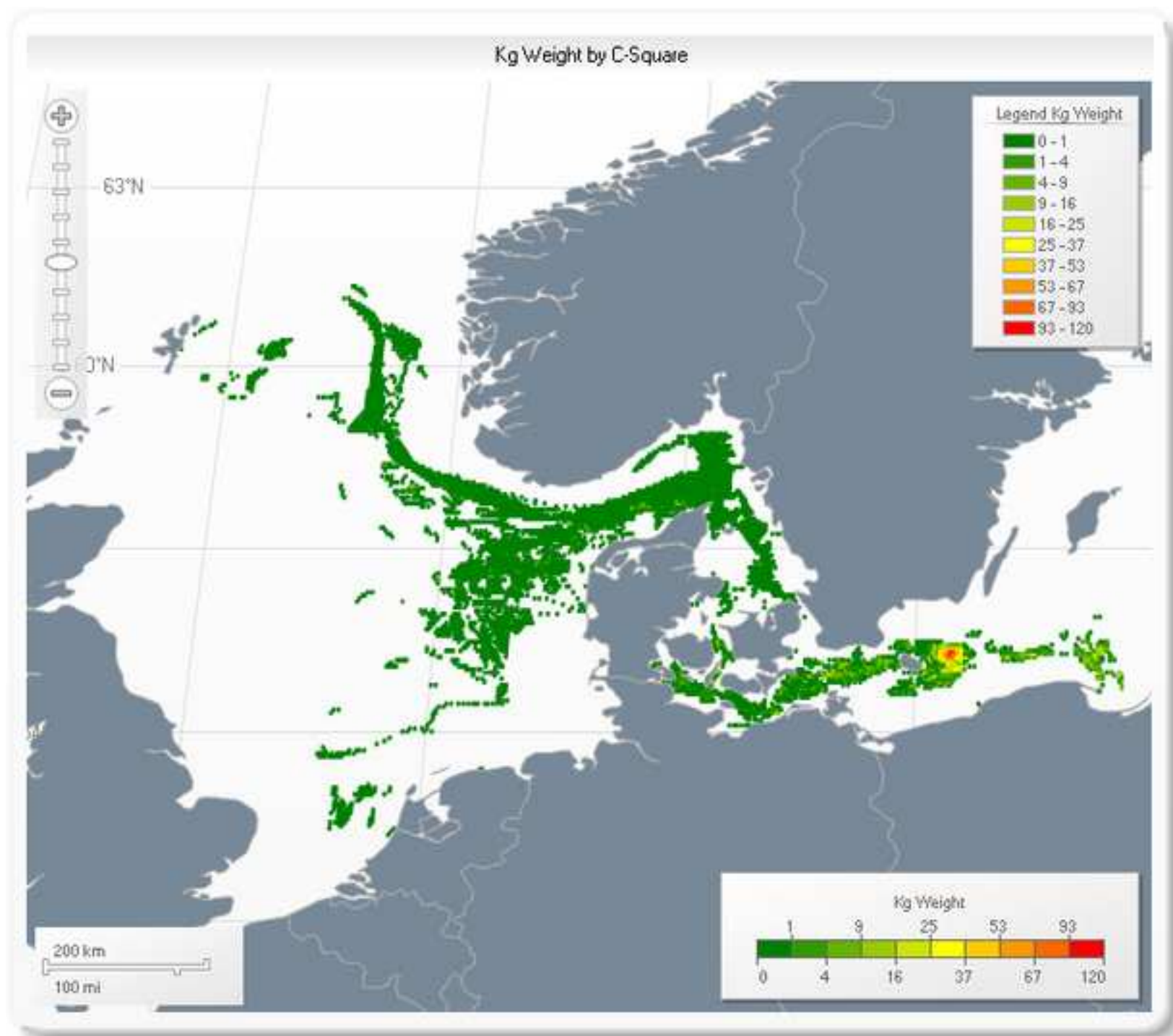


Figure 7 Color  
[Click here to download high resolution image](#)

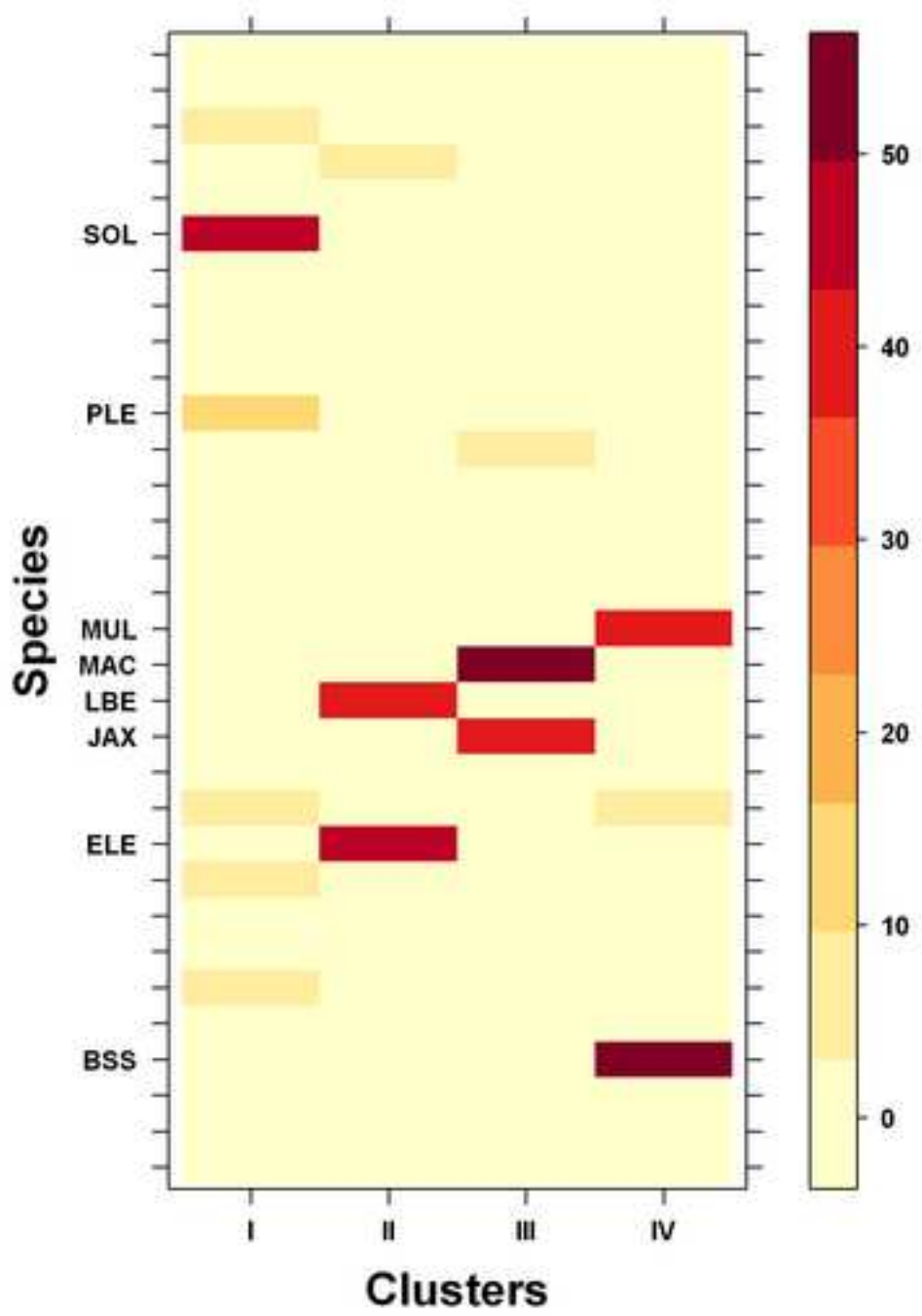




Figure 8 Color  
[Click here to download high resolution image](#)

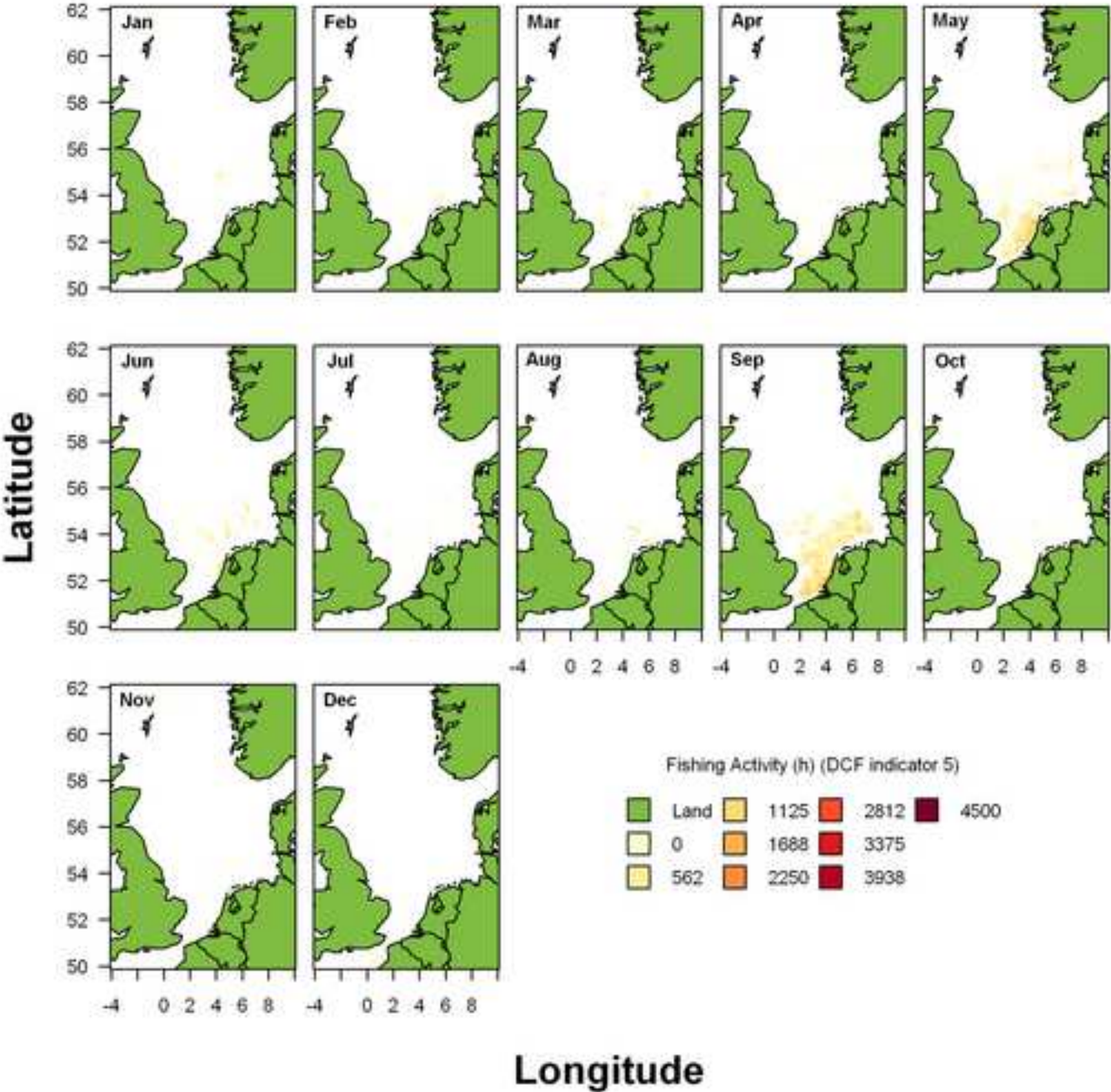


Figure 1 BW

[Click here to download high resolution image](#)

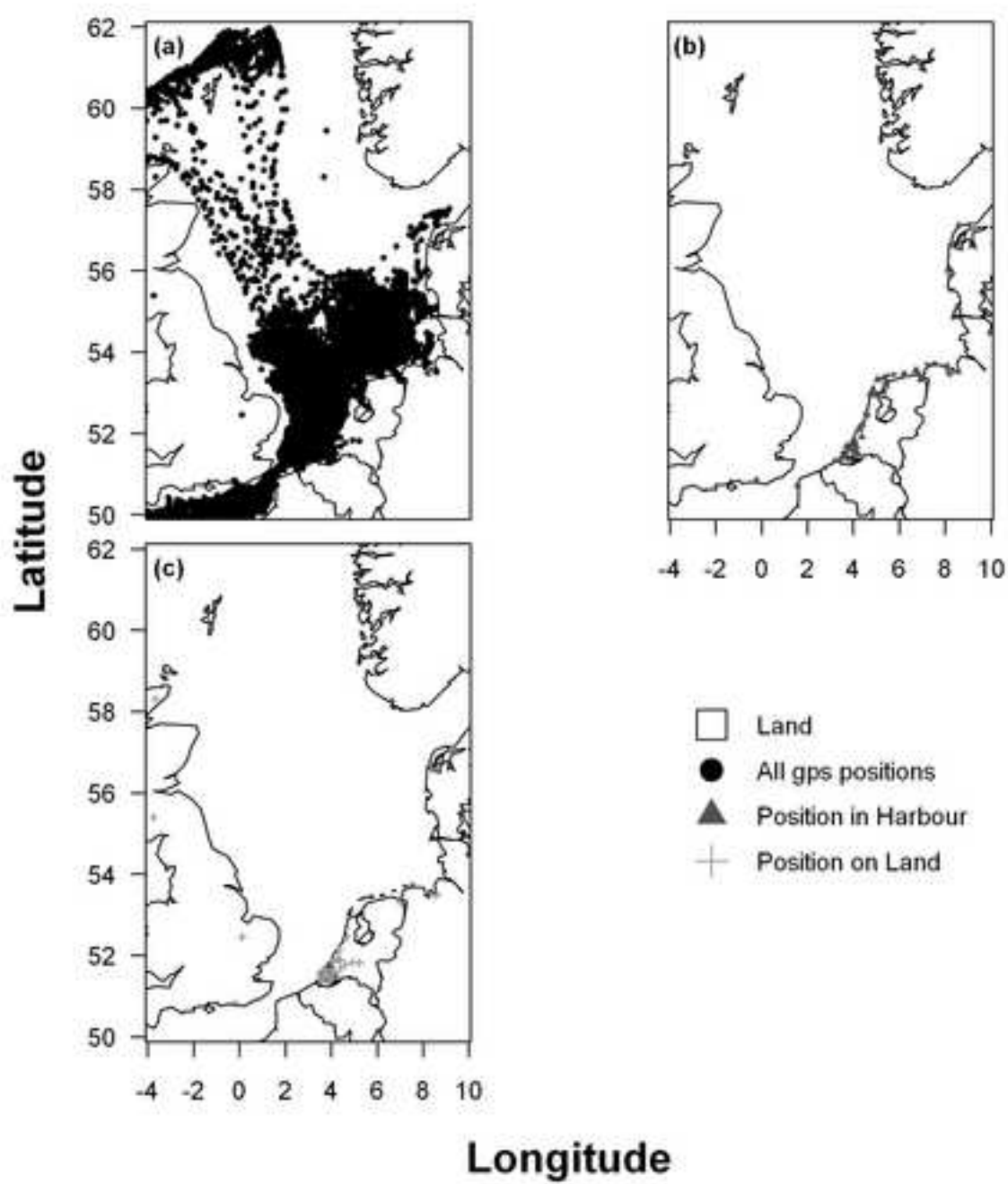


Figure 2 BW  
[Click here to download high resolution image](#)

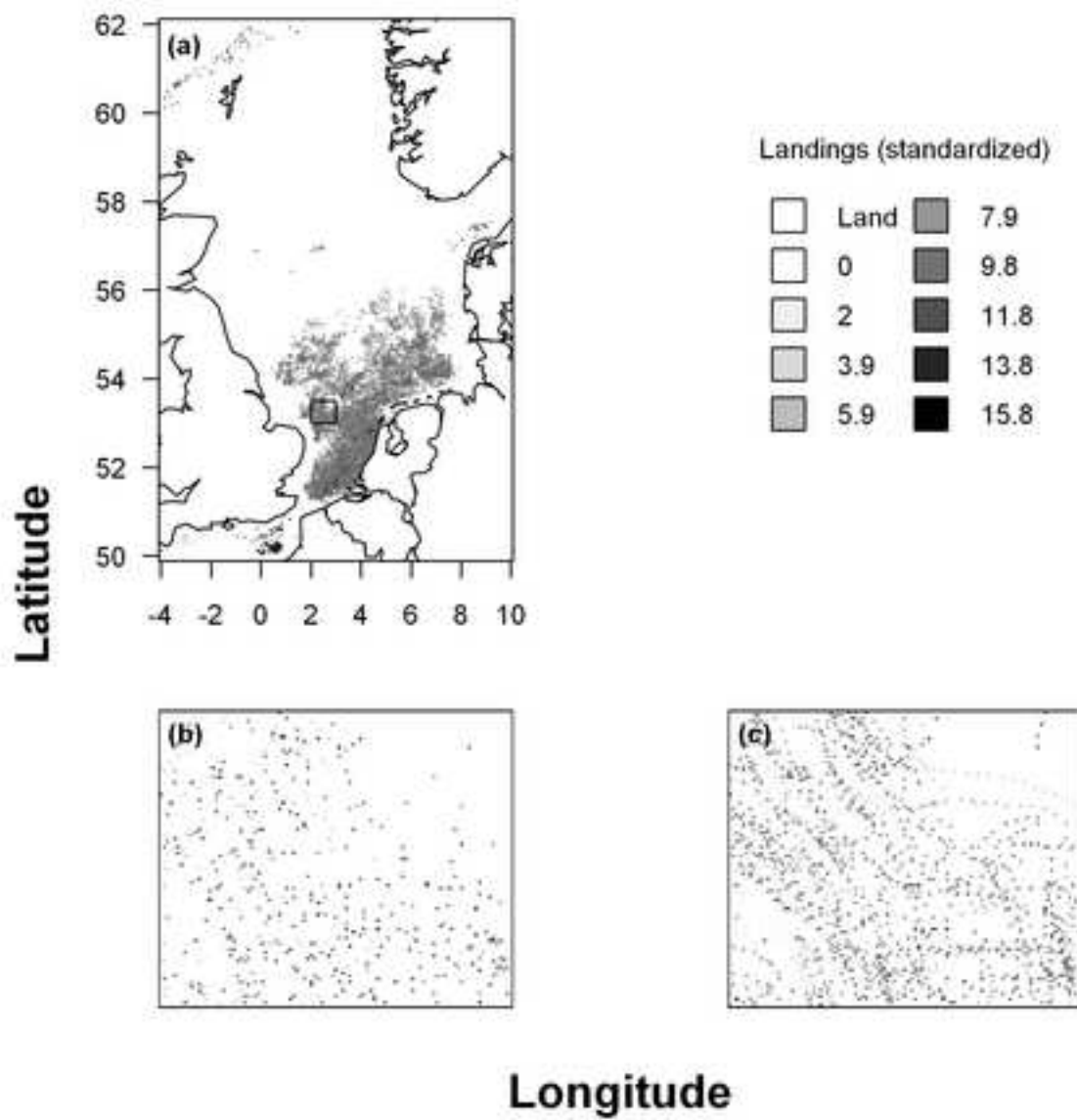




Figure 3 BW  
[Click here to download high resolution image](#)

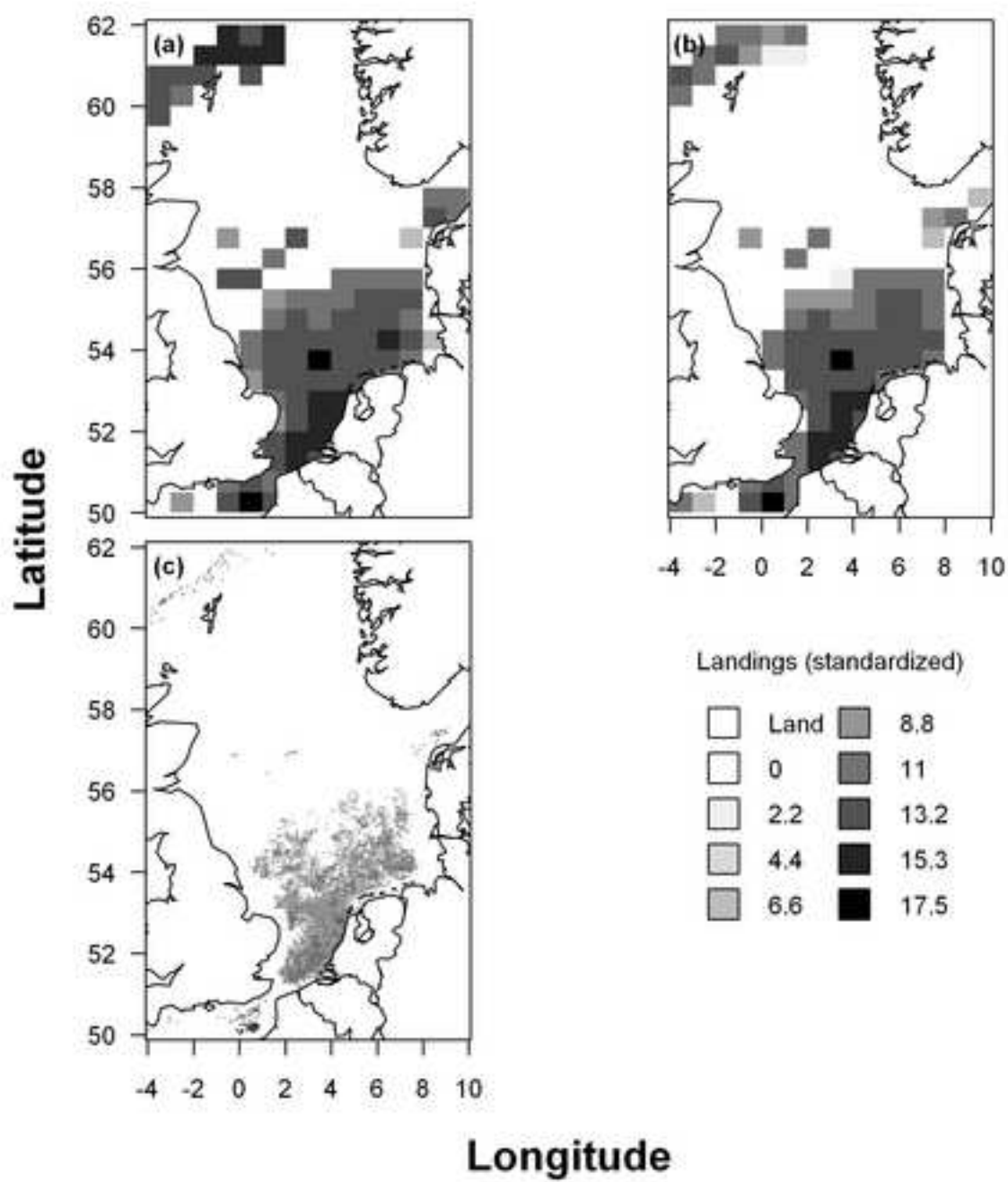


Figure 4 BW

[Click here to download high resolution image](#)

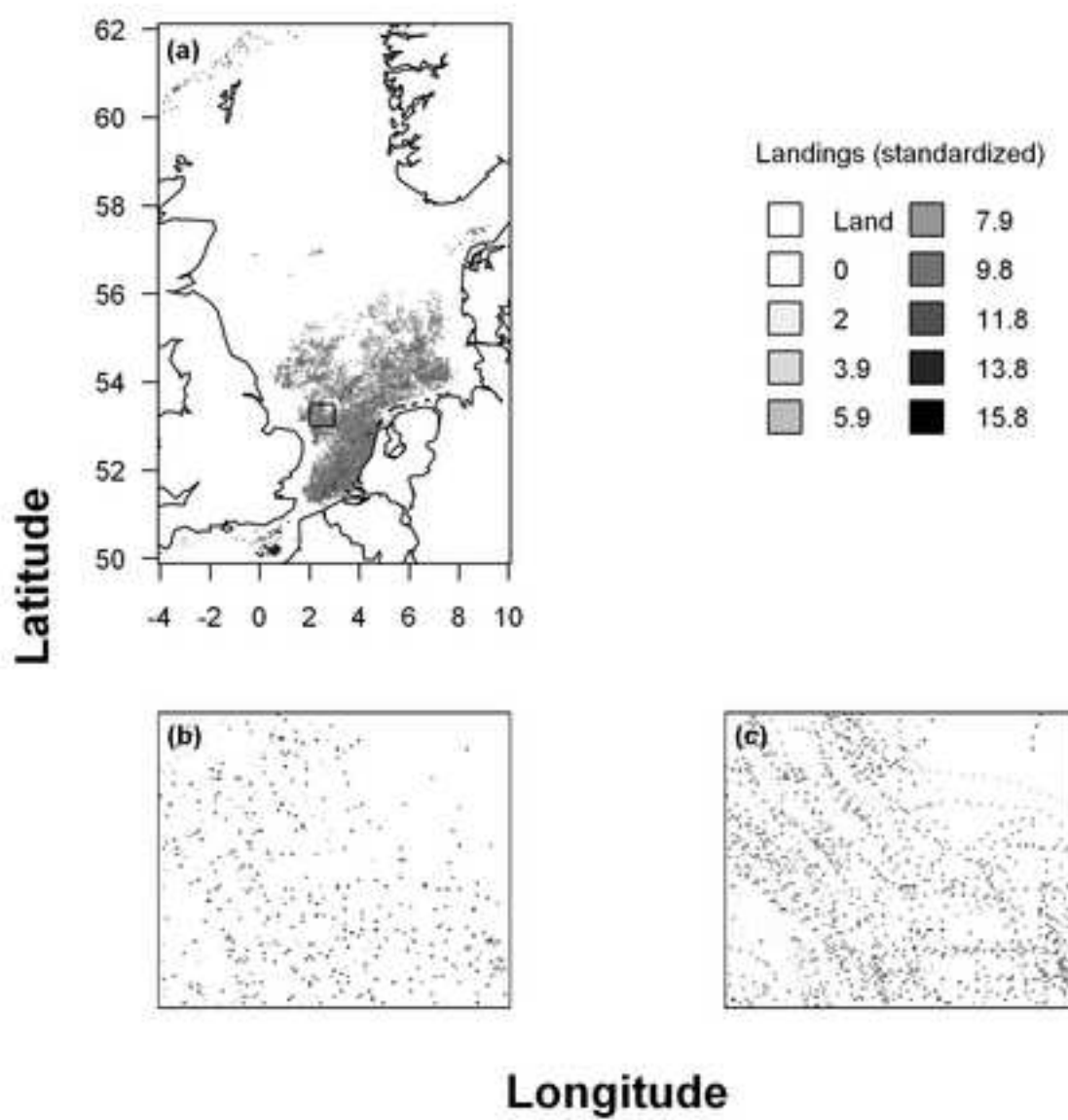


Figure 5 BW

[Click here to download high resolution image](#)

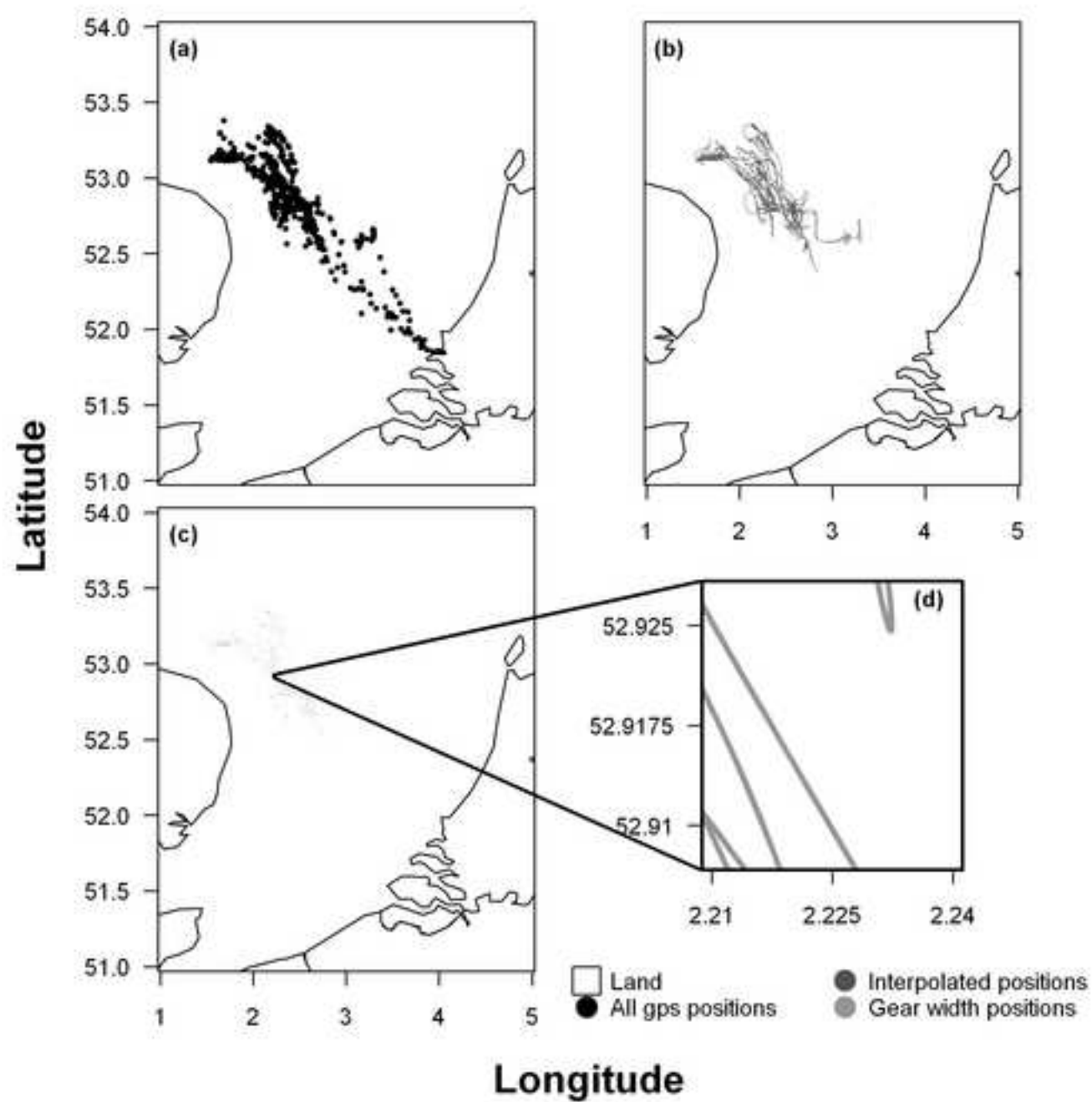


Figure 6 BW

[Click here to download high resolution image](#)

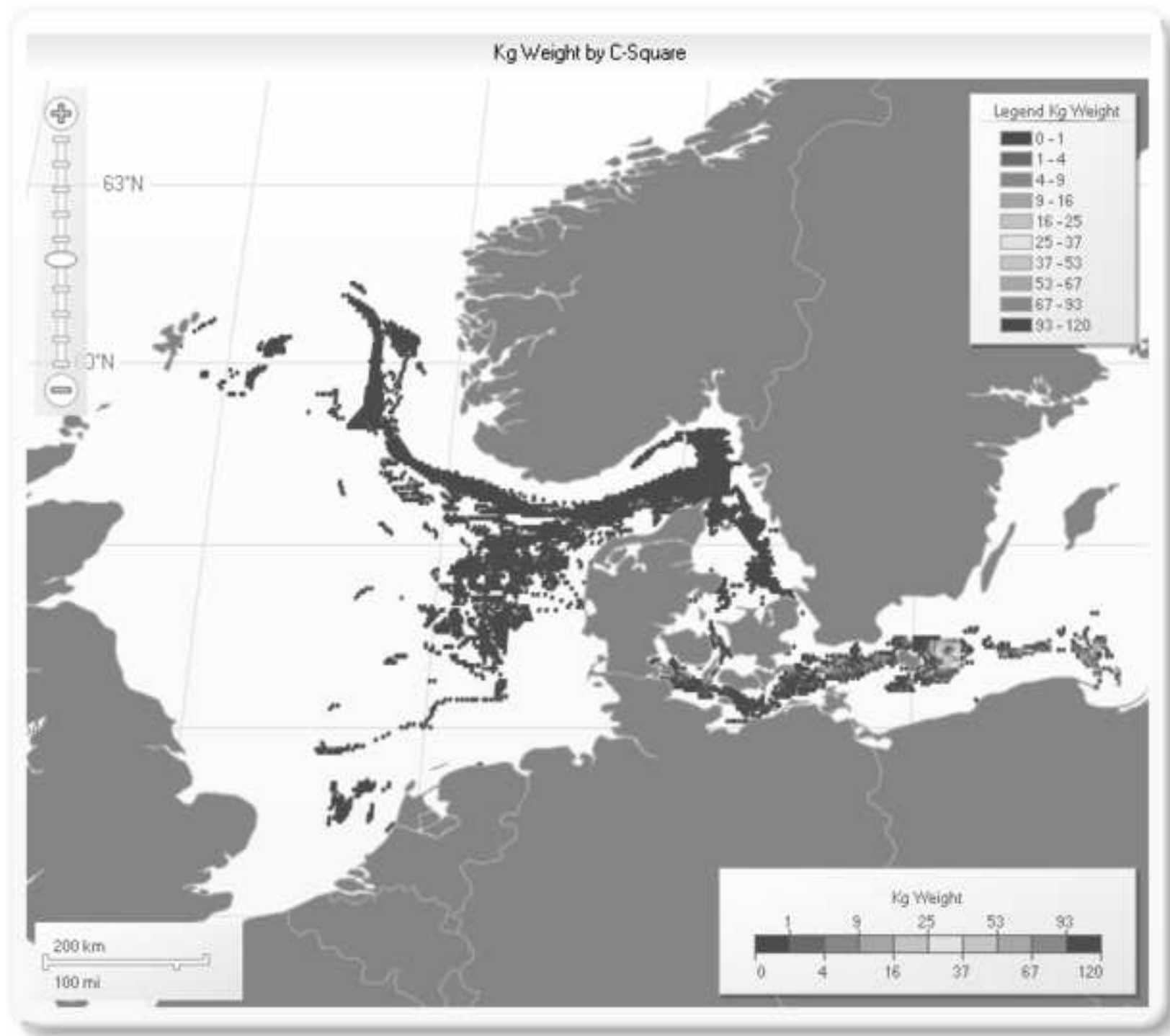


Figure 7 BW

[Click here to download high resolution image](#)

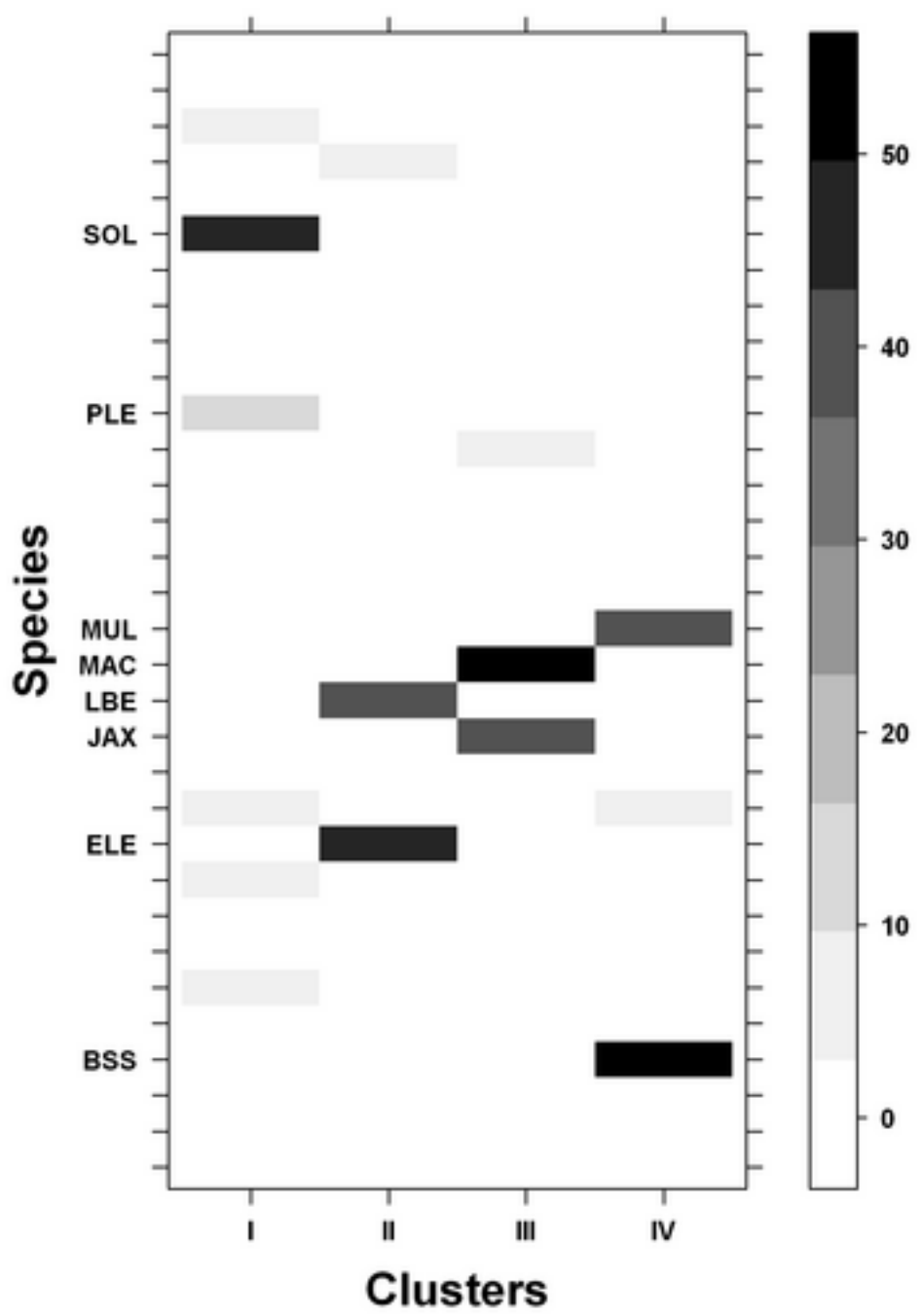
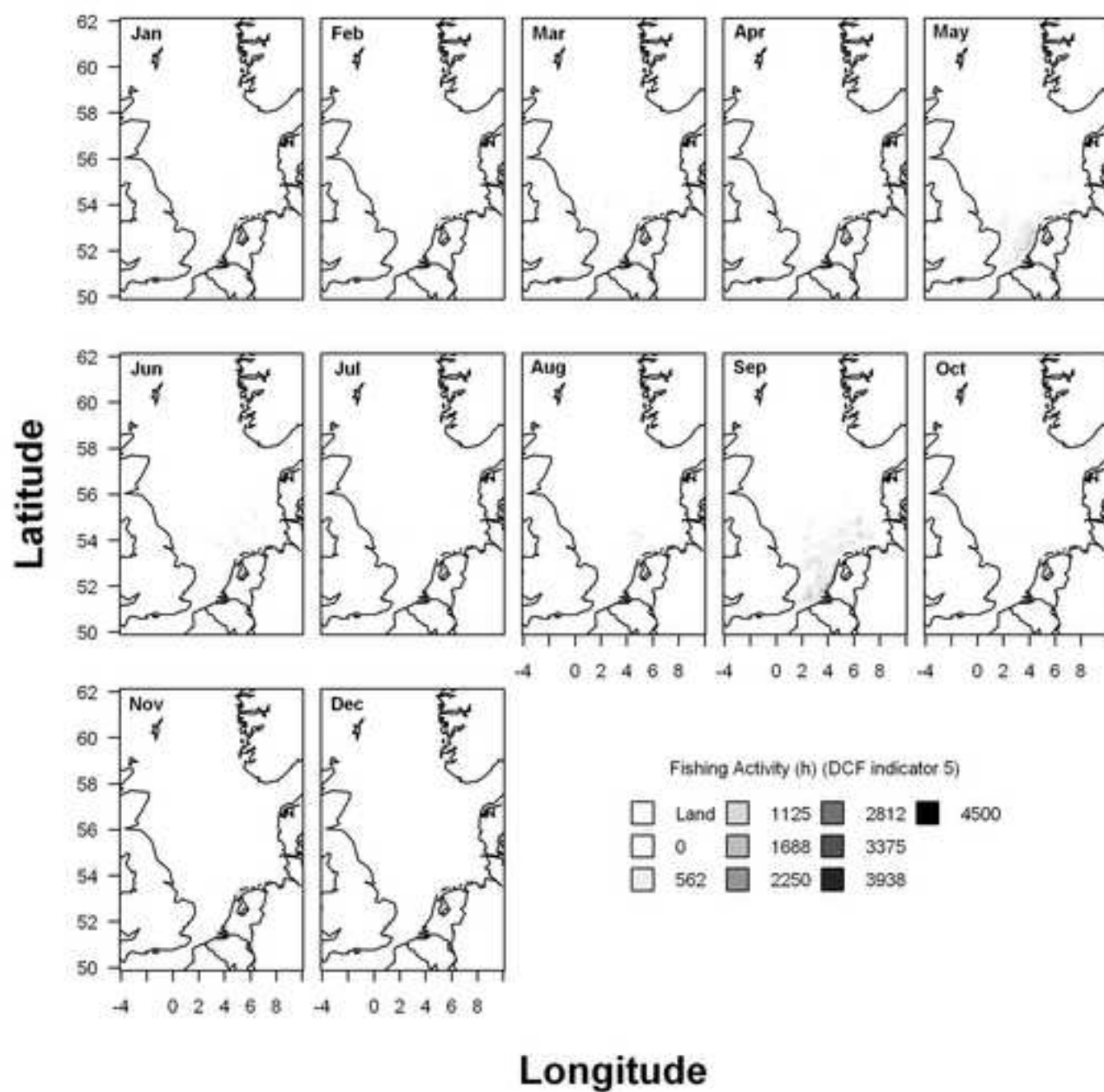


Figure 8 BW

[Click here to download high resolution image](#)



**Supplementary material for on-line publication only**

**[Click here to download Supplementary material for on-line publication only: vmstools\\_appendix\\_FR.doc](#)**